An improved estimator for non-Gaussianity in cosmic microwave background observations

Tristan L. Smith[1], Daniel Grin[2], and Marc Kamionkowski[3]

[1]*Berkeley Center for Cosmological Physics & Berkeley Lab University of California, Berkeley, CA 94720, USA*
[2]*School of Natural Sciences, Institute for Advanced Study, Princeton, New Jersey 08540, USA and*
[3]*Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD 21218, USA*

An improved estimator for the amplitude $f_{\rm NL}$ of local-type non-Gaussianity from the cosmic microwave background (CMB) bispectrum is discussed. The standard estimator is constructed to be optimal in the zero-signal (i.e., Gaussian) limit. When applied to CMB maps which have a detectable level of non-Gaussianity the standard estimator is no longer optimal, possibly limiting the sensitivity of future observations to a non-Gaussian signal. Previous studies have proposed an improved estimator by using a realization-dependent normalization. Under the approximations of a flat sky and a vanishingly thin last-scattering surface, these studies showed that the variance of this improved estimator can be significantly smaller than the variance of the standard estimator when applied to non-Gaussian CMB maps. Here this technique is generalized to the full sky and to include the full radiation transfer function, yielding expressions for the improved estimator that can be directly applied to CMB maps. The ability of this estimator to reduce the variance as compared to the standard estimator in the face of a significant non-Gaussian signal is re-assessed using the full CMB transfer function. As a result of the late time integrated Sachs-Wolfe effect, the performance of the improved estimator is degraded. If CMB maps are first cleaned of the late-time ISW effect using a tracer of foreground structure, such as a galaxy survey or a measurement of CMB weak lensing, the new estimator *does* remove a majority of the excess variance, allowing a higher significance detection of $f_{\rm NL}$.

I. INTRODUCTION

Over the past two decades our understanding of the physics of the early universe has gone from speculative to precise. We have numerous data sets which probe both the overall expansion of the universe as well as the statistics of the large-scale structure we see today. In addition to these observations, our standard cosmological model, with only six free parameters, provides a good fit to all of these data sets (see, e.g. [1]). The standard cosmological model relies on some basic assumptions about the origin and evolution of the universe: namely that soon after the big bang, the universe underwent a period of cosmic inflation during which nearly Gaussian perturbations were produced in an otherwise isotropic and homogeneous universe. After this period the universe was 'reheated'– i.e., populated with a thermal plasma of standard model particles. Particle physics dictates the interactions between the constituents of this primordial fluid, while general relativity dictates how the perturbations grow to form the structures we observe in both the clustering of galaxies, as well as in the anisotropies of the cosmic microwave background (CMB).

With increasingly precise observations, we may test the foundations of the standard cosmological model. Here we will be concerned with testing the assumption that the statistics of CMB fluctuations are Gaussian. Although small levels of non-Gaussianity may develop through non-linearities in the standard cosmological model [2–6], significant departures from Gaussianity can only be explained by changes to the fundamental physics of the early universe. For example, a detection of primordial non-Gaussianity could yield information on the interactions of the field (or fields) that seed the primordial curvature perturbation. If this field is the one that drives inflation (the inflaton), the measurement could yield insight into the detailed physics of inflation. If the curvature perturbation is seeded by a field that is sub-dominant during inflation (as in the curvaton scenario [7–10]), the primordial fluctuations may also be significantly non-Gaussian. Some more exotic possibilities, such as a non-canonical kinetic term for the inflaton [11–13], spatially modulated reheating [14], and novel initial vacuum states for the fluctuations [15], could also be probed by a detection or limit to non-Gaussianity in the CMB.

There are an infinite number of ways in which the statistics of the CMB may be non-Gaussian, although the effective field theory approach shows that inflationary theories can only produce three forms for the non-Gaussian primordial three-point correlation function [11]. In this study we restrict our attention to the 'local model' of non-Gaussianity, in which the primordial curvature perturbation, Φ, can be written in terms of an auxiliary Gaussian field ϕ as [2, 16–19]

$$\Phi(\vec{x}) = \phi(\vec{x}) + f_{\rm NL} \left[\phi^2(\vec{x}) - \langle \phi^2(\vec{x}) \rangle \right], \tag{1}$$

where the amplitude $f_{\rm NL}$ parameterizes the level of non-Gaussianity. This model is particularly important because

if $f_{\rm NL} \neq 0$ were detected then *all* single-field slow-roll inflation models would be ruled out [20, 21]. In addition, the local-type non-Gaussianity is predicted by the curvaton model [8–10], and so a detection could provide support for, or constraints to, this model.

Although there are several ways to estimate the level of non-Gaussianity in the CMB, an estimate of the harmonic three-point function (known as the CMB bispectrum) is the most sensitive [18, 22]. Given the large number of modes in the bispectrum (after restrictions due to statistical isotropy there are $l_{\rm max}^5$ terms in the bispectrum, where $l_{\rm max}(\simeq \sqrt{N_{\rm pix}})$ is the maximum multipole measured by a given experiment) a full exploration of the likelihood surface is computationally prohibitive, and so attempts to constrain the level of non-Gaussianity are made through estimators which have been constructed to minimize variance under the null hypothesis– i.e., when applied to CMB maps which are purely Gaussian [23, 24].

A direct application of the minimum-variance null-hypothesis (MVNH) estimator of $f_{\rm NL}$ from the CMB bispectrum is also computationally expensive since the computation involves a very large number of terms ($\sim l_{\rm max}^5$). The Planck satellite [25] will measure $l_{\rm max} \sim 2500$ multipole moments so a blind application of this estimator will take $\sim 10^{15}$ operations to compute! The real computational expense is even higher, as thousands of simulations must be run to characterize the statistics of this estimator. For a special set of non-Gaussian models (of which the local model is one) fast Fourier transforms (FFTs) may be used to greatly reduce this computational expense [26, 27]. Using FFTs the estimator may then be evaluated with a computation time that scales as $l_{\rm max}^3$, considerably reducing the computational expense [26, 28]. CMB data are currently consistent with vanishing non-Gaussianity [29, 30]; analysis of the WMAP 7-year data yields the 2-σ constraint $-10 < f_{\rm NL} < 74$ [1].

If future data remain consistent with the null hypothesis, then the standard MVNH estimator should have the minimum variance. As the data improves, if $f_{\rm NL}$ is found to be significantly different from zero, then the standard MVNH estimator is non-optimal and there may be other estimators which have a smaller variance (and hence increased signal-to-noise, S/N). This is because the MVNH estimator was *constructed* to have the minimum variance when $f_{\rm NL} = 0$ and when applied to data with $f_{\rm NL} \neq 0$ the variance is no longer minimized. In particular, when applied to the local model, for $f_{\rm NL} \neq 0$ the MVNH estimator exhibits a variance which is proportional to $f_{\rm NL}^2$, leading to a saturation of the S/N of the estimator, even for large values of $l_{\rm max}$ [31–33]. This indicates that a new method to estimate $f_{\rm NL}$ may extract a higher S/N from the measurements.

In one approach, estimators are abandoned, and the full likelihood surface is explored using Bayesian methods [34]. Although this approach has been shown to give improved constraints on $f_{\rm NL}$, it is computationally expensive, taking over 150,000 CPU hours to compute a best fit-value and confidence interval for $f_{\rm NL}$ with $l_{\rm max} = 512$. In another approach, an improved estimator is constructed by defining a new realization-normalized estimator (RNE) [31, 33]. This method normalizes the MVNH estimator using a particular combination of observed multipoles which are chosen so as to divide out the extra variance present. In Ref. [31] the RNE was derived in the flat sky and Sachs-Wolfe limit (in which the temperature anisotropies are given by fluctuations in the gravitational potential at the surface of last scattering, $\Delta T/T = -\Phi/3$ [35]). In these limits, it was shown in Ref. [31] that the RNE successfully removes all the variance proportional to $f_{\rm NL}^2$ of the MVNH estimator when $f_{\rm NL} \neq 0$.

In this work we generalize the RNE to include the full radiation transfer function and provide expressions that are valid on all-sky CMB maps. We compute the variance of the RNE and find that, unlike in the Sachs-Wolfe limit, only $\sim 50\%$ of the $f_{\rm NL}$-dependent variance is removed.

The residual $\mathcal{O}(f_{\rm NL}^2)$ variance is a result of the late-time, large-scale integrated Sachs-Wolfe (ISW) effect, in which CMB anisotropies are generated by gravitational potential decay along the line-of-sight [36–42]. To understand why the late-time ISW effect reduces our ability to remove this extra variance, we note that the form of the local-model non-Gaussianities implies that the bispectrum signal is dominated by harmonic-space triangles which correlate one large-scale mode with two small scale modes (i.e., 'squeezed' triangles). If the large-scale anisotropies are generated only by the Sachs-Wolfe effect then they can be inverted to give a direct measurement of the value of the large-scale gravitational potential at the surface of last scattering. These measurements, in turn, allow us to directly infer the non-Gaussian contribution to the large-scale anisotropies, leading to a complete removal of all of the excess variance when $f_{\rm NL} \neq 0$. However, in the presence of *both* the late-time ISW and Sachs-Wolfe effect, this inversion is not possible, leading to a degradation of the performance of the RNE. We find, however, that if a tracer of foreground structure (such as a wide-field galaxy survey or the deflection field responsible for weak lensing of the CMB) is first applied to 'clean' a CMB map of the late-time ISW effect then $\sim 80 - 90\%$ of the $f_{\rm NL}$-dependent variance may be removed by using the RNE.

We begin by summarizing the non-Gaussian local model and its bispectrum in Sec. II. We provide an intuitive, geometric argument for the origin of the $f_{\rm NL}$-dependent variance that afflicts the local-model MVNH estimator in Sec. III. The realization-dependent normalization (RNE) is derived in Sec. IV using a method that applies in the presence of the full radiation transfer function. In Sec. V we compute the variance of the RNE using the full radiation transfer function, finding that the $f_{\rm NL}$-dependent variance of the MVNH estimator is only partially removed. In Sec. VI, we show that the $f_{\rm NL}$-dependent variance is further reduced by first cleaning the CMB of the late-time

ISW contribution using a large-scale structure survey [such as the NRAO VLA Sky Survey (NVSS), or the next-generation Joint Dark Energy Mission (WFIRST) proposal] or a measurement of CMB lensing. Conventions for flat-sky approximations employed in numerical calculations are stated in Appendix A. Detailed expressions needed to obtain the variance are derived in Appendix B. In Appendix C we present a computationally efficient algorithm (using fast Fourier transforms) to compute our estimator on full-sky CMB maps.

II. THE NON-GAUSSIAN LOCAL MODEL

First we will review the basic equations that relate to the definition and estimation of the CMB bispectrum, restricting attention to the local model as defined in Eq. (1). The CMB bispectrum is defined by

$$B_{l_1 l_2 l_3}^{m_1 m_2 m_3} \equiv \langle a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} \rangle = \mathcal{G}_{l_1 l_2 l_3}^{m_1 m_2 m_3} b_{l_1 l_2 l_3}, \tag{2}$$

where the a_{lm} are the usual multipole moments of the temperature map, $b_{l_1 l_2 l_3}$ is the reduced CMB bispectrum, and the Gaunt integral is given by

$$\mathcal{G}_{l_1 l_2 l_3}^{m_1 m_2 m_3} \equiv \int d^2 \hat{n} Y_{l_1 m_1}(\hat{n}) Y_{l_2 m_2}(\hat{n}) Y_{l_3 m_3}(\hat{n}) = \sqrt{\frac{(2l_1+1)(2l_2+1)(2l_3+1)}{4\pi}} \begin{pmatrix} l_1 & l_2 & l_2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \tag{3}$$

and $\begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix}$ is a Wigner $3J$-coefficient. A product of three multipoles form an unbiased estimator for the angle-averaged CMB bispectrum:

$$B_{l_1 l_2 l_3}^{\rm obs} = \sum_{m_1 m_2 m_3} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} \tag{4}$$

The minimum-variance null-hypothesis (MVNH) estimator for $f_{\rm NL}$ is given by [23, 24]

$$\widehat{f_{\rm NL}} = \sigma_0^{-2} \sum_{l_1 \leq l_2 \leq l_3} \frac{B_{l_1 l_2 l_3}^{\rm obs} B_{l_1 l_2 l_3}}{C_{l_1} C_{l_2} C_{l_3}}, \tag{5}$$

where

$$\langle a_{lm} a_{l'm'}^* \rangle = C_l \delta_{ll'} \delta_{mm'}, \tag{6}$$

and $B_{l_1 l_2 l_3}$ is the primordial (i.e., theoretical) angle-averaged bispectrum. The normalization of this estimator is given by the variance under the null hypothesis, σ_0^2 [43]:

$$\sigma_0^2 = \sum_{l_1 \leq l_2 \leq l_3} \frac{B_{l_1 l_2 l_3}^2}{\Delta_{l_1 l_2 l_3} C_{l_1} C_{l_1} C_{l_3}}, \tag{7}$$

where $\Delta_{l_1 l_2 l_3} = 1$ if $l_1 \neq l_2 \neq l_3$, 6 if $l_1 = l_2 = l_3$, and 2 otherwise. Finally, the angle-averaged primordial bispectrum, $B_{l_1 l_2 l_3}$, is given in terms of the reduced bispectrum

$$B_{l_1 l_2 l_3} = \sqrt{\frac{(2l_1+1)(2l_2+1)(2l_3+1)}{4\pi}} \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix} b_{l_1 l_2 l_3}. \tag{8}$$

Now restricting attention to the local-model bispectrum, at any radial location along the line of sight the primordial curvature potential, Φ, can be decomposed into spherical harmonics. The local model ansatz in Eq. (1) then implies that

$$\Phi_{l_1 m_1}(r) = \phi_{l_1 m_1}(r) + f_{\rm NL}(-1)^{m_1} \sum_{l_2 l_3} \sum_{m_2 m_3} \mathcal{G}_{l_1 l_2 l_3}^{-m_1 m_2 m_3} \phi_{l_2 m_2}(r) \phi_{l_3 m_3}(r). \tag{9}$$

This allows us to write the reduced bispectrum in a line-of-sight integral:

$$b_{l_1 l_2 l_3} = 2 \left(\int r^2 dr \alpha_{l_1}(r) \beta_{l_2}(r) \beta_{l_3}(r) + \text{cyclic permutations.} \right), \tag{10}$$

where the two filter-functions are given in terms of the transfer functions

$$\alpha_l(r) \equiv \frac{2}{\pi} \int k^2 dk j_l(kr) S_l(k), \tag{11}$$

$$\beta_l(r) \equiv \frac{2}{\pi} \int k^2 dk j_l(kr) S_l(k) P(k), \tag{12}$$

and $S_l(k)$ is the CMB temperature anisotropy transfer function [44]. The generalization to polarization is straightforward [43]; in this work we limit our attention to the signature of primordial non-Gaussianity in the CMB temperature.

Finally, we will need expressions for various auto and cross-correlations in terms of the distance along the line-of-sight r. The power-spectrum for the auxiliary Gaussian field ϕ is

$$\left\langle \phi(\vec{k})\phi^*(\vec{k}') \right\rangle = (2\pi)^3 \delta^{(3)}(\vec{k} - \vec{k}') P(k), \tag{13}$$

where

$$\phi(\vec{x}) = \int \frac{d^3\vec{k}}{(2\pi)^3} \phi(\vec{k}) e^{i\vec{k}\cdot\vec{x}}. \tag{14}$$

These expressions lead to the line-of-sight autocorrelation [45]

$$\begin{aligned} \langle \phi_{lm}(r)\phi^*_{l'm'}(r') \rangle &= \frac{2}{\pi} \delta_{l,l'} \delta_{m,m'} \int k^2 dk P(k) j_l(kr) j_l(kr') \\ &\equiv \chi(r,r') \delta_{l,l'} \delta_{m,m'}. \end{aligned} \tag{15}$$

Similarly, the temperature multipole moments can be written as [26, 46]

$$a_{lm} = \int r^2 dr \alpha_l(r) \Phi_{lm}(r). \tag{16}$$

With this we can write the line-of-sight cross-correlation

$$\begin{aligned} \langle a_{lm}\phi^*_{l'm'}(r) \rangle &= \frac{2}{\pi} \delta_{ll'} \delta_{mm'} \int k^2 dk P(k) j_l(kr) S_l(k) \\ &\equiv \delta_{ll'} \delta_{mm'} \beta_l(r). \end{aligned} \tag{17}$$

These expressions will be useful in the following discussion.

III. THE ORIGIN OF THE INCREASED VARIANCE

The standard non-Gaussian estimator written in Eq. (5) is only optimal under the null hypothesis: $f_{\rm NL} = 0$. In the case where $f_{\rm NL} \neq 0$ the estimator becomes suboptimal (in the sense that it no longer saturates minimum variance attainable according to the Cramer-Rao bound, see Refs. [23, 31]) and has a variance which depends on $f_{\rm NL}$. Specifically, in the flat-sky Sachs-Wolfe limit the variance scales as [31, 33, 47]

$$\begin{aligned} \left\langle (\Delta \widehat{f_{\rm NL}})^2 \right\rangle &= \sigma_0^2 + f_{\rm NL}^2 \sigma_1^2, \\ &= \frac{1}{72 A f_{\rm sky} l_{\rm max}^2 \ln(l_{\rm max})} + f_{\rm NL}^2 \frac{1}{2 \ln^3(l_{\rm max})}, \end{aligned} \tag{18}$$

where A is the amplitude of the power-spectrum of the primordial curvature perturbations, $P(k) = Ak^{-3}$. From this expression we can see that for $f_{\rm NL} \neq 0$ that in the large $l_{\rm max}$ limit the variance flattens out and scales as $\ln^{-3}(l_{\rm max})$.

To understand the origin of this $f_{\rm NL}$-dependent variance let us first consider a simple toy model. Let a_i be a random variable with $\langle a_i \rangle = 0$ and $\langle a_i a_j \rangle = \sigma^2 \delta_{ij}$. As we will discuss further, the $f_{\rm NL}$-dependent variance in the $\widehat{f_{\rm NL}}$ estimator comes from terms which look like

$$\hat{X} \equiv \sum_{ij} a_i a_j. \tag{19}$$

The mean and variance of $\hat{X}$ are

$$\left\langle \hat{X} \right\rangle = \sum_{ij} \langle a_i a_j \rangle = N\sigma^2, \tag{20}$$

$$\left\langle (\Delta \hat{X})^2 \right\rangle = \sum_{ijkl} \langle a_i a_j a_k a_l \rangle - \sigma^4 \delta_{ij} \delta_{kl} = 2N^2 \sigma^4 \tag{21}$$

where N is the number of data points. From this, it is clear that the S/N of $\hat{X}$ is given by

$$\frac{S}{N} = \frac{\left\langle \hat{X} \right\rangle}{\sqrt{\left\langle (\Delta \hat{X})^2 \right\rangle}} = \frac{1}{\sqrt{2}}. \tag{22}$$

We can see that the S/N is a constant. The reason is simple: since the a_i are uncorrelated, the off diagonal terms in $\hat{X}$ with $i \neq j$ do not contribute to the signal; on the other hand, they *do* contribute to the noise, leading to a constant S/N[1]. We will see that something similar happens for the MVNH estimator.

To understand the origin of the increased variance we expand the MVNH estimator in Eq. (5) in powers of $f_{\rm NL}$

$$\widehat{f_{\rm NL}} \approx \mathcal{B}_0 + f_{\rm NL} \mathcal{B}_1, \tag{23}$$

with

$$\mathcal{B}_0 = \sigma_0^{-2} \sum_{l_1 \leq l_2 \leq l_3} \sum_{m_1 m_2 m_3} \frac{a^{\rm L}_{l_1 m_1} a^{\rm L}_{l_2 m_2} a^{\rm L}_{l_3 m_3}}{C_{l_1} C_{l_2} C_{l_3}} B_{l_1 l_2 l_3} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix}, \tag{24}$$

$$\mathcal{B}_1 = \sigma_0^{-2} \sum_{l_1 \leq l_2 \leq l_3} \sum_{m_1 m_2 m_3} \frac{a^{\rm NL}_{l_1 m_1} a^{\rm L}_{l_2 m_2} a^{\rm L}_{l_3 m_3}}{C_{l_1} C_{l_2} C_{l_3}} B_{l_1 l_2 l_3} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} + \text{cyclic permutations}, \tag{25}$$

where

$$a^{\rm L}_{lm} = \int r^2 dr \alpha_l(r) \phi_{lm}(r), \tag{26}$$

$$a^{\rm NL}_{lm} = (-1)^m \sum_{l_1 l_2} \sum_{m_1 m_2} \mathcal{G}^{-m m_1 m_2}_{l l_1 l_2} \int r^2 dr \alpha_l(r) \phi_{l_1 m_1}(r) \phi_{l_2 m_2}(r). \tag{27}$$

The total observed multipole moment is

$$a_{lm} \simeq a^{\rm L}_{lm} + f_{\rm NL} a^{\rm NL}_{lm}. \tag{28}$$

We now shed light on the origin of the increased variance by considering the flat-sky and Sachs-Wolfe limit of Eq. (25). In this limit the radiation transfer function is given by $S_l(k) = -j_l(kr_*)/3$ where r_* is the conformal distance to the surface of last scattering. We refer the reader to Appendix A for expressions which relate the full-sky to flat-sky expressions. In these limits we have [31, 33]:

$$C_l = \frac{A}{9\pi l(l+1)}, \tag{29}$$

$$\mathcal{B}_1 = 3\sigma_0^2 \sum_{\vec{m}, \vec{n}, \vec{l}_2, \vec{l}_3} \frac{b_{l_1 l_2 l_3}}{2\Omega^3 C_{l_1} C_{l_2} C_{l_3}} \left\{ a_{\vec{m}} a_{\vec{n}} \delta_{\vec{m}+\vec{n},\vec{l}_1} \right\} \left\{ a_{\vec{l}_2} a_{\vec{l}_3} \delta_{\vec{l}_2+\vec{l}_3,-\vec{l}_1} \right\}, \tag{30}$$

where the primordial power-spectrum is taken to be scale invariant, $P(k) = A/k^3$. Rewriting $\mathcal{B}_1$ in this way we can see that it is a *product of two triangles with a shared side* $\vec{l}_1$. We show a graphical representation of this in the top section of Fig. 1. We can rewrite $\mathcal{B}_1$ schematically as

$$\mathcal{B}_1 = \sum_{i,j} W(i) A_i A_j, \tag{31}$$

―――――

[1] Of course, if we were interested in estimating the variance we would choose a different weighting than in Eq. (19) and use the estimator $\sum_i a_i^2$ which has the same signal but a reduced variance compared to Eq. (19). As we will discuss further, the increased variance in the $\widehat{f_{\rm NL}}$ estimator cannot be reduced by choosing a new weighting.

with i labeling the triangle $i = \{\vec{l}_1, \vec{l}_2, \vec{l}_3\}$ and j labeling the triangle $j = \{\vec{l}_1, \vec{m}, \vec{n}\}$.
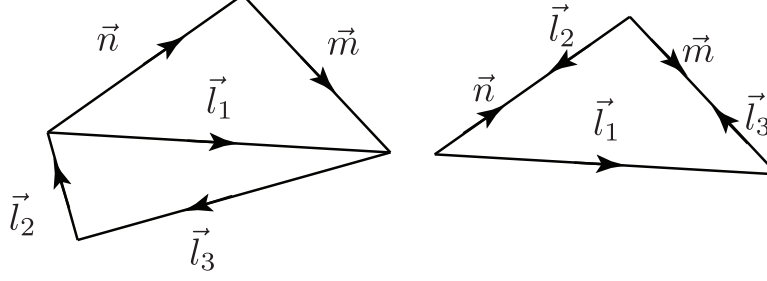

FIG. 1: A graphical representation of the temperature configurations that are included in the $\mathcal{B}_1$ part of the bispectrum estimator, Eq. (25). The full shape is composed of two distinct triangles. Only one triangle contributes to the mean, but all triangles contribute to the variance, leading to a S/N which decreases slowly with $l_{\max}$.

The mean is given by a sum over a single triangle, as shown in the right panel of Fig. 1:

$$\langle \mathcal{B}_1 \rangle = \sum_{i,j} W(i) \langle A_i A_j \rangle = \sum_i W(i) \langle A_i^2 \rangle . \tag{32}$$

Therefore, we can see that the mean is composed of a sum of $(l_{\max}^2)^2 = l_{\max}^4$ terms. On the other hand, Fig. 1 shows that the variance of $\mathcal{B}_1$ is computed from a sum of a product of two triangles which gives $(l_{\max}^2)^4 = l_{\max}^8$. As in the simple example given above, this shows that there are a large number of terms which do not contribute to the mean of $\mathcal{B}_1$ but *do* contribute to the variance, leading to a signal-to-noise (S/N) which does not decrease as fast as one might have expected with increasing $l_{\max}$.

It might seem that this can be corrected by choosing a different weight function $W(i) = W_{\vec{l}_1, \vec{l}_2, \vec{l}_3}$. Since this weight was chosen to optimize the estimator under the null (i.e., $f_{\mathrm{NL}} = 0$) hypothesis it is natural to think that a different weighting will be needed when $f_{\mathrm{NL}} \neq 0$. Unfortunately, since the terms in $\mathcal{B}_1$ which contribute to the variance but not the mean involve contractions between $(\vec{l}_1, \vec{l}_2)$ and $(\vec{n}, \vec{m})$, it is clear that *no choice of weight*, which depends only on the 'observed' indices $(\vec{l}_1, \vec{l}_2, \vec{l}_3)$, will remove these terms. Therefore, a new weighting cannot decrease the variance of this estimator in the high S/N regime. Another way to remove the terms which do not contribute to the signal but do contribute to the variance is to use a realization dependent normalization.

Before we discuss how to construct a realization dependent normalization that reduces the variance in $\mathcal{B}_1$ let us first demonstrate in what sense the squeezed limit dominates the variance in $\mathcal{B}_1$. In the flat-sky Sachs-Wolfe limits, it has been shown that the variance is well approximated by [33]

$$\langle (\Delta \mathcal{B}_1)^2 \rangle \propto \sigma_0^4 \sum_{\vec{l}_1 + \vec{l}_2 + \vec{l}_3 = 0} \frac{b_{l_1 l_2 l_3}}{C_{l_1}} \sum_{\vec{s}_1 + \vec{s}_2 = \vec{l}_3} \frac{b_{s_1 s_2 s_3}}{C_{s_1}} \delta_{\vec{l}_3 + \vec{s}_3, 0} . \tag{33}$$

where $\{\vec{l}\}$ indicates the sum is over $\vec{l}_1 + \vec{l}_2 + \vec{l}_3 = 0$. We can calculate this by first writing it in a less compact form: In these limits the bispectrum is given by [23]

$$b_{l_1 l_2 l_3} = -6\{C_{l_2} C_{l_3} + C_{l_3} C_{l_1} + C_{l_1} C_{l_2}\}. \tag{34}$$

The second sum in Eq. (33) is given by

$$\sum_{\vec{s}_1 + \vec{s}_2 = \vec{l}_3} \frac{b_{s_1 s_2 s_3}}{C_{s_1}} \delta_{\vec{l}_3 + \vec{s}_3, 0} = -\frac{2}{3\pi} \sum_{\vec{s}_1 + \vec{s}_2 = \vec{l}_3} \left[\frac{s_1(s_1 + 1)}{s_2(s_2 + 1)l_3(l_3 + 1)} + \frac{1}{l_3(l_3 + 1)} + \frac{1}{s_2(s_2 + 1)} \right] \tag{35}$$

and the summand is dominated by triangles where one index is much less than the other two (i.e., the squeezed limit) so that

$$\sum_{\vec{s}_1 + \vec{s}_2 = \vec{l}_3} \frac{b_{s_1 s_2 s_3}}{C_{s_1}} \propto A \left(\frac{l_{\max}}{l_3} \right)^2 . \tag{36}$$

We then have that

$$\langle(\Delta\mathcal{B}_1)^2\rangle \propto \sigma_0^4 l_{\mathrm{max}}^2 A^2 \sum_{\vec{l}_1+\vec{l}_2+\vec{l}_3=0} \frac{l_1^2+l_2^2+l_3^2}{l_2^2 l_3^4}. \tag{37}$$

We can see from this expression that it will be dominated by the terms where l_3 is smallest. Therefore, taking the squeezed limit again we write $l_3 \ll l_2 \simeq l_1$ and find that

$$\langle(\Delta\mathcal{B}_1)^2\rangle \propto \sigma_0^4 A^2 \frac{l_{\mathrm{max}}^4}{l_{\mathrm{min}}^2} \propto \ln(l_{\mathrm{max}})^{-2}, \tag{38}$$

where we have taken the limit $l_{\mathrm{max}} \gg l_{\mathrm{min}}$ and used the expression for σ_0 found in Eq. (18).

The numerical calculation found in Ref. [33] shows that the variance decreases slightly faster with $\langle(\Delta\mathcal{B}_1)^2\rangle \propto \ln(l_{\mathrm{max}})^{-3}$ - non-squeezed configurations do make some contribution to the estimator which causes the variance fall off more rapidly with l_{max}. For other forms of non-Gaussianity, the $\mathcal{O}(f_{\mathrm{NL}})$ term in the expansion of the estimator can not be written simply as in Eq. (25). An expression analogous to Eq. (30), however, may still be written though with a different weighting due to the shapes of the Fourier-space triangles which dominate the S/N in these cases. Since these models are not dominated by the squeezed limit, they are not afflicted by the slow scaling with l_{max} of the S/N that occurs for the local model.

IV. REALIZATION NORMALIZED ESTIMATOR

In order to improve the f_{NL} estimator when $f_{\mathrm{NL}} \neq 0$, we will search for a realization dependent normalization which removes some of the variance in those terms in the estimator which do not contribute to the signal but do contribute to the noise. To do this we follow the approach presented in Ref. [31] and construct a new realization-dependent normalization which will remove as much of the f_{NL}-dependent variance of $\mathcal{B}_1$ as possible. In other words, we wish to define a realization-dependent normalization that is highly correlated with $\mathcal{B}_1$.

We write the new realization-dependent normalization as an estimator for $\mathcal{B}_1$, $\widehat{\mathcal{B}_1}$, so that we obtain a new estimator for f_{NL}, the realization-normalized estimator (RNE):

$$(\widehat{f_{\mathrm{NL}}})^{\mathrm{N}} = \frac{\widehat{f_{\mathrm{NL}}}}{\widehat{\mathcal{B}_1}} \approx \frac{\mathcal{B}_0}{\widehat{\mathcal{B}_1}} + f_{\mathrm{NL}}\frac{\mathcal{B}_1}{\widehat{\mathcal{B}_1}}. \tag{39}$$

The extent to which the new realization-dependent normalization decreases the variance of this estimator is determined by the correlation between $\mathcal{B}_1$ and $\widehat{\mathcal{B}_1}$; in the limit that these two terms are fully correlated the f_{NL}^2 variance is completely removed.

Looking at the equation for $\mathcal{B}_1$ [Eq. (25)] the only term that cannot be immediately written in terms of observables is a_{lm}^{NL}. Therefore, when constructing the estimator $\widehat{\mathcal{B}_1}$ we must find a minimum variance estimator for a_{lm}^{NL}. To do this we define a weighted sum,

$$\widehat{a}_{lm}^{\mathrm{NL}} = \sum_{l_1 l_2} \sum_{m_1 m_2} W_{ll_1l_2}^{mm_1m_2} a_{l_1m_1} a_{l_2m_2}, \tag{40}$$

and demand that the weight minimizes the variance,

$$\frac{\delta}{\delta W_{ll_1l_2}^{mm_1m_2*}} \left\langle \left| \widehat{a}_{lm}^{\mathrm{NL}} - a_{lm}^{\mathrm{NL}} \right|^2 \right\rangle = 0. \tag{41}$$

Choosing the weight

$$W_{ll_1l_2}^{mm_1m_2} = \frac{\mathcal{G}_{ll_1l_2}^{mm_1m_2}}{C_{l_1}C_{l_2}} \int r^2 dr \alpha_l(r)\beta_{l_1}(r)\beta_{l_2}(r), \tag{42}$$

satisfies Eq. (41) and thus leads to a minimum variance estimator for a_{lm}^{NL}. With this, the estimator for $\mathcal{B}_1$ can be written

$$\widehat{\mathcal{B}_1} = \sigma_0^{-2} \sum_{l_1 \leq l_2 \leq l_3} \sum_{m_1m_2m_3} \frac{\widehat{a}_{lm}^{\mathrm{NL}} a_{l_1m_1} a_{l_2m_2}}{C_l C_{l_1} C_{l_2}} B_{ll_1l_2} \begin{pmatrix} l & l_1 & l_2 \\ m & m_1 & m_2 \end{pmatrix} + \text{cyclic permutations}, \tag{43}$$

$$= \sigma_0^{-2} \sum_{l_1 \leq l_2 \leq l_3, l_a l_b} \sum_{m_1m_2m_3m_am_b} \frac{a_{l_1m_1} a_{l_2m_2} a_{l_am_a} a_{l_bm_b}}{2C_l C_{l_1} C_{l_2} C_{l_a} C_{l_b}} B_{ll_1l_2} B_{ll_al_b} \begin{pmatrix} l & l_1 & l_2 \\ m & m_1 & m_2 \end{pmatrix} \begin{pmatrix} l & l_a & l_b \\ m & m_a & m_b \end{pmatrix}, \tag{44}$$

where we have made the approximation $a_{lm}^{\mathrm{L}} \simeq a_{lm}$, which is true to lowest order in f_{NL}.

We note that there are several other ways of arriving at the same estimator for $\mathcal{B}_1$. In particular, if we were to instead search for a minimum variance estimator for Φ_{lm} as in Ref. [26], then we would again be lead to the estimator for $\mathcal{B}_1$ in Eq. (44). Furthermore, the same relation between the underlying gravitational potential and the observed multipoles appears when calculating the shape of the likelihood surface for the MVNH estimator as discussed in Ref. [24].

As we now discuss, the improved estimator given by Eqs. (39) and (44) (the RNE), has two important properties. First we show that in the case where the transfer function is just the Sachs-Wolfe transfer function, the RNE is the same as the estimator presented in Ref. [31]. Our discussion then shows how the RNE, in this limit, is able to remove all of the f_{NL}-dependent variance from the standard MVNH estimator. Second, when the full transfer-function is used, the combination of both the Sachs-Wolfe and late-time integrated Sachs-Wolfe (ISW) effects on large angular scales causes the RNE to be less effective; it only removes part of the f_{NL}-dependent variance. Using other tracers of large-scale structure, as discussed in Ref. [48], allows us to improve the performance of the RNE, so that it can remove *most* of the f_{NL}-dependent variance. Finally, in Appendix C we discuss how our estimator can be rewritten in terms of real-space quantities, so as to be computationally efficient.

V. PROPERTIES OF THE REALIZATION-NORMALIZED ESTIMATOR

To simplify the calculations in this section we work in the flat-sky approximation. Although the results presented will differ when compared to full-sky calculations, since we are interested in calculating the *fractional* reduction of the f_{NL}-dependent variance (relative to the f_{NL}-dependent variance of the standard MVNH estimator), we expect the differences to be small.

To quantify the statistical properties of the RNE we must calculate the mean and variance of the ratios $\mathcal{B}_0/\widehat{(\mathcal{B}_1)}$ and $\mathcal{B}_1/\widehat{(\mathcal{B}_1)}$. To do this we will use an approximate formula for the variance of the ratio of two stochastic variables. Let x_1 and x_2 be two stochastic variables with means μ_1 and μ_2, variances σ_1^2 and σ_2^2, and covariance ρ. We wish to calculate the mean and variance of

$$W = \frac{x_1}{x_2}. \tag{45}$$

The probability density function (PDF) of W may be computed analytically if x_1 and x_2 are normally distributed [49]. However, the formula for the PDF is quite complicated and it does not immediately yield analytic formulae for the mean, $\langle W \rangle$, and variance $\langle (W - \langle W \rangle)^2 \rangle \equiv \langle \Delta W^2 \rangle$. Instead, we will derive an approximate expression. To do this we write $x_1 = \mu_1 + \delta x_1$ and $x_2 = \mu_2 + \delta x_2$. We then assume $\delta x_{1,2}/\mu_{1,2} = \epsilon \ll 1$; this will be true near the peak of the normal distribution if $\mu_{1,2} \gg \sigma_{1,2}$ (we have verified that the stochastic quantities discussed we are interested in satisfy this condition). We can write

$$W \approx \frac{\mu_1}{\mu_2} \left(1 + \delta x_1/\mu_1 - \delta x_2/\mu_2 \right) + \mathcal{O}(\epsilon^2). \tag{46}$$

With this we can easily compute

$$\langle W \rangle = \frac{\mu_1}{\mu_2}, \tag{47}$$

$$\langle \Delta W^2 \rangle = \frac{\sigma_1^2}{\mu_1^2} + \frac{\sigma_2^2}{\mu_2^2} - 2\frac{\rho}{\mu_1 \mu_2}. \tag{48}$$

A. The mean

We wish to show that $\langle \mathcal{B}_1 \rangle = \langle \widehat{\mathcal{B}_1} \rangle$. To leading order in f_{NL} in the flat-sky approximation,

$$\langle \mathcal{B}_1 \rangle = \sigma_0^2 \sum_{\vec{l}_1 + \vec{l}_2 + \vec{l}_3 = 0} \delta_{\vec{k} + \vec{k}', \vec{l}_1} \frac{B(l_1, l_2, l_3)}{2\Omega^2 C_{l_1} C_{l_2}} \int r^2 dr \, \alpha_{l_1}(r) \left\langle a_{\vec{L}}^{\mathrm{L}} a_{\vec{l}_2}^{\mathrm{L}} \phi_{\vec{k}}(r) \phi_{\vec{k}'}(r) \right\rangle, \tag{49}$$

$$\left\langle \widehat{\mathcal{B}_1} \right\rangle = \sigma_0^2 \sum_{\vec{l}_1 + \vec{l}_2 + \vec{l}_3 = 0} \delta_{\vec{k} + \vec{k}', \vec{l}_1} \frac{B(l_1, l_2, l_3)}{2\Omega^2 C_{l_1} C_{l_2}} \int r^2 dr \, \alpha_{l_1}(r) \frac{\beta_k(r)\beta_{k'}(r)}{C_k C_{k'}} \left\langle a_{\vec{L}}^{\mathrm{L}} a_{\vec{l}_2}^{\mathrm{L}} a_{\vec{k}}^{\mathrm{L}} a_{\vec{k}'}^{\mathrm{L}} \right\rangle. \tag{50}$$

The fact that $|\vec{l}_1| \geq 2$ requires that the Wick contraction which contains $\langle \phi_{\vec{k}}(r)\phi_{\vec{k}'}(r)\rangle$ does not contribute to the mean. Therefore we are left with

$$\left\langle a_{\vec{l}_2} a_{\vec{L}} \phi_{\vec{k}}(r)\phi_{\vec{k}'}(r)\right\rangle = \left\langle a_{\vec{l}_2}\phi_{\vec{k}}(r)\right\rangle\left\langle a_{\vec{L}}\phi_{\vec{k}'}(r)\right\rangle + \left\langle a_{\vec{L}}\phi_{\vec{k}}(r)\right\rangle\left\langle a_{\vec{l}_2}\phi_{\vec{k}'}(r)\right\rangle, \tag{51}$$

$$= \beta_{l_2}(r)\beta_L(r)\left[\delta_{\vec{l}_2,-\vec{k}}\delta_{\vec{L},-\vec{k}'} + \delta_{\vec{l}_2,-\vec{k}'}\delta_{\vec{L},-\vec{k}}\right]. \tag{52}$$

Now to calculate the equivalent expression for $\widehat{\mathcal{B}}_1$. Again, as with $\mathcal{B}_1$ only the 'cross' terms survive and we have

$$\frac{\beta_k(r)\beta_{k'}(r)}{C_k C_{k'}}\left\langle a_{\vec{l}_2} a_{\vec{L}} a_{\vec{k}} a_{\vec{k}'}\right\rangle = \beta_{l_2}(r)\beta_L(r)\left[\delta_{\vec{l}_2,-\vec{k}}\delta_{\vec{L},-\vec{k}'} + \delta_{\vec{l}_2,-\vec{k}'}\delta_{\vec{L},-\vec{k}}\right], \tag{53}$$

so we have that $\langle\mathcal{B}_1\rangle = \left\langle\widehat{\mathcal{B}}_1\right\rangle$. With this we can conclude that the RNE is unbiased.

B. The variance

The variance of the first term can be approximated by

$$\left\langle\left(\Delta\frac{(\mathcal{B}_0)}{(\widehat{\mathcal{B}}_1)}\right)^2\right\rangle \approx \langle(\mathcal{B}_0)^2\rangle + \left\langle(\mathcal{B}_0)^2\Delta(\widehat{\mathcal{B}_1})\Delta(\widehat{\mathcal{B}_1})\right\rangle \tag{54}$$

$$\approx \langle(\mathcal{B}_0)(\mathcal{B}_0)\rangle\left(1 + \left\langle\Delta(\widehat{\mathcal{B}_1})\Delta(\widehat{\mathcal{B}_1})\right\rangle\right), \tag{55}$$

$$= \sigma_0^2\left(1 + \left\langle\Delta(\widehat{\mathcal{B}_1})\Delta(\widehat{\mathcal{B}_1})\right\rangle\right). \tag{56}$$

The variance of the second term can be written as

$$\left\langle\left(\Delta\frac{(\mathcal{B}_1)}{(\widehat{\mathcal{B}}_1)}\right)^2\right\rangle \approx \langle\Delta(\mathcal{B}_1)\Delta(\mathcal{B}_1)\rangle + \left\langle\Delta(\widehat{\mathcal{B}_1})\Delta(\widehat{\mathcal{B}_1})\right\rangle - 2\left\langle\Delta(\mathcal{B}_1)\Delta(\widehat{\mathcal{B}_1})\right\rangle. \tag{57}$$

A tedious yet straight-forward calculation shows that the variance of the RNE is composed of four terms (we show the details of the calculation in Appendix A). Of those four terms one dominates in the $l_{\max} \gg 1$ limit so that we have

$$\langle\Delta(\mathcal{B}_1)\Delta(\mathcal{B}_1)\rangle = 8\sigma_0^4 \sum_{\{\vec{l}\},\{\vec{k}\}} \frac{B(l_1,l_2,l_3)B(k_1,k_2,k_3)}{C_{l_1}C_{l_2}C_{k_1}C_{k_2}C_{k_3}}\delta_{\vec{l}_3,\vec{k}_3} \tag{58}$$

$$\times \int r^2 dr (r')^2 dr' \alpha_{l_1}(r)\alpha_{k_1}(r')\beta_{k_2}(r')\beta_{l_2}(r)\chi_{l_3}(r,r'),$$

$$\left\langle\Delta(\widehat{\mathcal{B}_1})\Delta(\widehat{\mathcal{B}_1})\right\rangle = 8\sigma_0^4 \sum_{\{\vec{l}\},\{\vec{k}\}} \frac{B(l_1,l_2,l_3)B(k_1,k_2,k_3)}{C_{l_1}C_{l_2}C_{k_1}C_{k_2}C_{k_3}}\delta_{\vec{l}_3,\vec{k}_3} \tag{59}$$

$$\times \int r^2 dr (r')^2 dr' \alpha_{l_1}(r)\alpha_{k_1}(r')\beta_{k_2}(r')\beta_{l_2}(r)\frac{\beta_{l_3}(r)\beta_{l_3}(r')}{C_{l_3}},$$

$$\left\langle\Delta(\mathcal{B}_1)\Delta(\widehat{\mathcal{B}_1})\right\rangle = \langle\Delta(\mathcal{B}_1)\Delta(\mathcal{B}_1)\rangle. \tag{60}$$

The variance is given by

$$\left\langle\left(\Delta\frac{\mathcal{B}_1}{\widehat{\mathcal{B}}_1}\right)^2\right\rangle = 8\sigma_0^4 \sum_{\{\vec{l}\},\{\vec{k}\}} \frac{B(l_1,l_2,l_3)B(k_1,k_2,k_3)}{2C_{l_1}C_{l_2}C_{k_1}C_{k_2}C_{k_3}}\delta_{\vec{l}_3,\vec{k}_3}$$

$$\times \int r^2 dr (r')^2 dr' \alpha_{l_1}(r)\alpha_{k_1}(r')\beta_{l_2}(r)\beta_{k_2}(r')\mathcal{D}_{l_3}(r,r'), \tag{61}$$

$$\mathcal{D}_l(r,r') \equiv \left[\chi_l(r,r') - \frac{\beta_l(r)\beta_l(r')}{C_l}\right]. \tag{62}$$

We are interested in calculating the *fractional reduction* of the f_{NL}-dependent variance in the RNE, $(\widehat{f_{\mathrm{NL}}})^{\mathrm{N}}$, relative to the f_{NL}-dependent variance in the standard MVNH estimator, $\widehat{f_{\mathrm{NL}}}$. To quantify this we will define

$$\mathcal{R} \equiv \left\langle \left(\Delta \frac{\mathcal{B}_1}{\widehat{\mathcal{B}}_1} \right)^2 \right\rangle \bigg/ \langle \Delta \mathcal{B}_1^2 \rangle. \tag{63}$$

If $\mathcal{R} = 0$ then all of the f_{NL}-dependent variance has been removed; for $0 < \mathcal{R} < 1$ then there is a residual f_{NL}-dependent variance which the RNE does not remove.
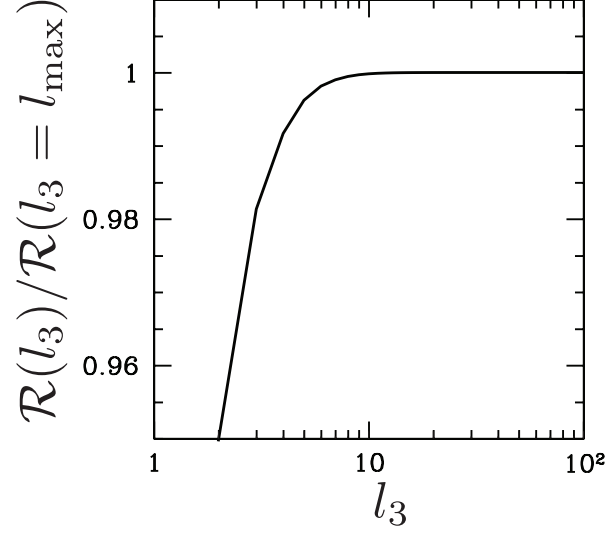
FIG. 2: The ratio of the fractional reduction of the f_{NL}-dependent variance in $(\widehat{f_{\mathrm{NL}}})^{\mathrm{N}}$ as a function of the value of the l_3 multipole (which is the large scale in the squeezed limit). This figure was produced with $l_{\max} = 100$ (we have checked that other choices for $l_{\max}$ reproduce curve). From this figure we can see that $\sim 95\%$ of the fractional reduction, $\mathcal{R}$, is contained in the quadrupole $l_3 = 2$ justifying our approximation in Eq. (65).

Looking at Eq. (61) we can see that the extent to which we reduce the variance of the RNE depends on the function $D_l(r,r')$. In the squeezed limit ($|\vec{l}_3| \sim l_{\min} \ll |\vec{l}_1| \simeq |\vec{l}_2|$ and $|\vec{k}_3| \sim l_{\min} \ll |\vec{k}_1| \simeq |\vec{k}_2|$), we have

$$\left\langle \left(\Delta \frac{\mathcal{B}_1}{\widehat{\mathcal{B}}_1} \right)^2 \right\rangle \propto \sum_{l_3} \mathcal{D}_{l_3}(r_*, r_*) \sum_{\{\vec{l}\}, \{\vec{k}\}} \frac{B(l_1, l_2, l_3) B(k_1, k_2, l_3)}{C_{l_1} C_{l_2} C_{l_3} C_{k_1} C_{k_2}} \beta_{l_2}(r_*) \beta_{k_2}(r_*), \tag{64}$$

where we have made the substitution $r^2 \alpha_l(r) \propto \delta(r - r_*)$ for $l \gg l_{\min}-$ i.e., to a very good approximation, all small-scale power is sourced by physics that occurs at the surface of last scattering. This allows us to write the fractional reduction of the f_{NL}-dependent variance as

$$\mathcal{R} \approx \frac{\mathcal{D}_2(r_*, r_*)}{\chi_2(r_*, r_*)}, \tag{65}$$

where we have fixed $l_3 = l_{\min} = 2$ since that value contains upwards of 95% of the variance (see Fig. 2). We have verified this approximation by calculating the full sum in Eq. (61) and found it to be accurate to a few percent. The function $\chi_l(r, r')$ is calculated using $P(k) = A k^{n_s - 4}$, where we use $n_s = 1$ to simplify the computation of the integral over k (the dependence on the actual value of n_s is negligible).

C. The f_{NL}-dependent variance in the Sachs-Wolfe limit

Before calculating the improvement in the f_{NL}-dependent variance in the case of a ΛCDM universe, let us first analyze these expressions in the Sachs-Wolfe limit. In this limit $\alpha_l(r) \propto \delta(r - r_*)$. From Eq. (62) it is clear that the f_{NL}-dependent variance will vanish (i.e., $\mathcal{R} = 0$) since it depends on

$$\mathcal{D}_l(r_*, r_*) = \chi_l(r_*, r_*) - \beta_l^2(r_*)/C_l = 0. \tag{66}$$

This result was found in Refs. [21, 31, 32].

We can understand this result in a slightly different way which will highlight how the Sachs-Wolfe limit is unique. In the Sachs-Wolfe limit the estimator $\widehat{a}_{lm}^{\rm NL}$ is *fully correlated* (up to order $Af_{\rm NL}^2$) with $a_{lm}^{\rm NL}$ since

$$\widehat{a}_{lm}^{\rm NL} = -3 \sum_{l_1 l_2} \sum_{m_1 m_2} (-1)^m \mathcal{G}_{ll_1 l_2}^{-mm_1 m_2} a_{l_1 m_1} a_{l_2 m_2}, \tag{67}$$

$$= -\frac{1}{3} \sum_{l_1 l_2} \sum_{m_1 m_2} (-1)^m \mathcal{G}_{ll_1 l_2}^{-mm_1 m_2} \phi_{l_1 m_1}(r_*) \phi_{l_2 m_2}(r_*) + \mathcal{O}(Af_{\rm NL}^2) \tag{68}$$

$$a_{lm}^{\rm NL} = \sum_{l_1 l_2} \sum_{m_1 m_2} (-1)^m \mathcal{G}_{ll_1 l_2}^{-mm_1 m_2} \int r^2 dr \alpha_l(r) \phi_{l_1 m_1}(r) \phi_{l_2 m_2}(r) \tag{69}$$

$$= -\frac{1}{3} \sum_{l_1 l_2} \sum_{m_1 m_2} (-1)^m \mathcal{G}_{ll_1 l_2}^{-mm_1 m_2} \phi_{l_1 m_1}(r_*) \phi_{l_2 m_2}(r_*). \tag{70}$$

In a ΛCDM universe, however, the large-scale anisotropies receive power from *both* the Sachs-Wolfe as well as the late-time integrated Sachs-Wolfe (ISW) effects. This additional contribution to the large-scale power degrades the correlation between $\widehat{a}_{lm}^{\rm NL}$ and $a_{lm}^{\rm NL}$ thus leaving some residual $f_{\rm NL}$-dependent variance.

D. The $f_{\rm NL}$-dependent variance for the full ΛCDM transfer function

We are now in a position to calculate the reduction of the $f_{\rm NL}$-dependent variance using the realization-dependent normalization in a ΛCDM universe. As we already discussed, the value of the ratio $\mathcal{D}_l(r_*, r_*)/\chi_l(r_*, r_*)$ at the quadrupole ($l = 2$) gives a good estimate for the fractional reduction of the $f_{\rm NL}$-dependent variance, $\mathcal{R}$.
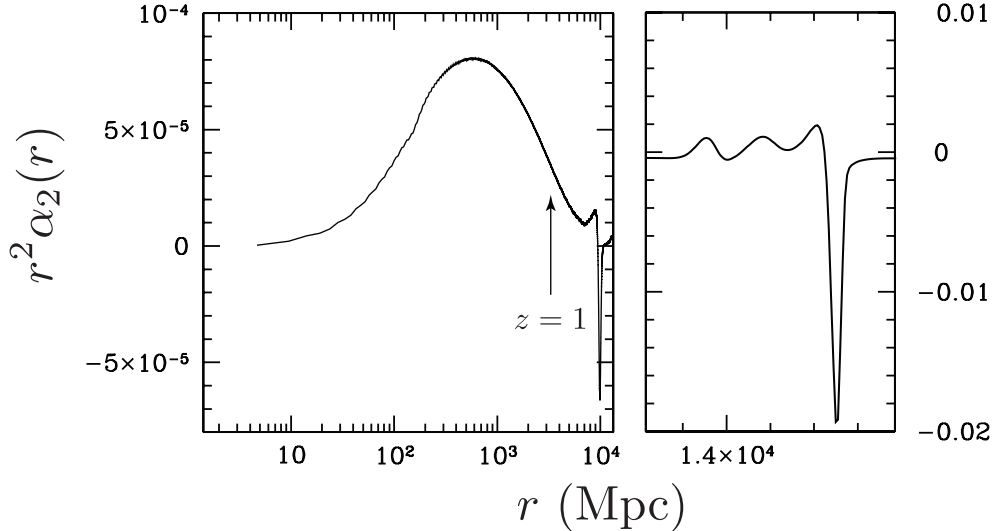

FIG. 3: The transfer function, $r^2 \alpha_l(r)$, for the quadrupole ($l = 2$) as a function of conformal distance in Mpc. The left-hand panel shows the effects of late-time ISW. The dip around $r \approx 10^4$ Mpc corresponds to reionization at $z = 11$. The right-hand panel highlights the evolution of $r^2 \alpha_l(r)$ around decoupling. Note that the scale is not the same in both panels.

In Fig. 3 we show how the quadrupole filter $\alpha_2(r)$ [see Eq. (11)] depends on the line-of-sight distance r in a ΛCDM universe. First note that there is a relatively large rise below $r \leq 10^3$ Mpc, corresponding to the late-time ISW effect. Furthermore at $r \simeq 2 \times 10^4$ Mpc, the filter shows a sharp dip corresponding to the contribution from the surface of last scattering. Therefore in a ΛCDM universe on large scales the observed multipoles a_{lm} have contributions from two separate epochs: the ISW effect at late-times and the Sachs-Wolfe effect around the surface of last scattering. This two-epoch contribution is central to understanding how the RNE works when applied to a ΛDCM universe.

To investigate how the late-time ISW effect impacts the improvement of the RNE, we modified the publicly available Boltzmann code CAMB [50] to include a new parameter, $A_{\rm ISW}$, which controls the amplitude of the late-time ISW effect in the line-of-sight integral [44]. When $A_{\rm ISW} = 1$ it takes on its standard value in the calculation of the C_ls; when
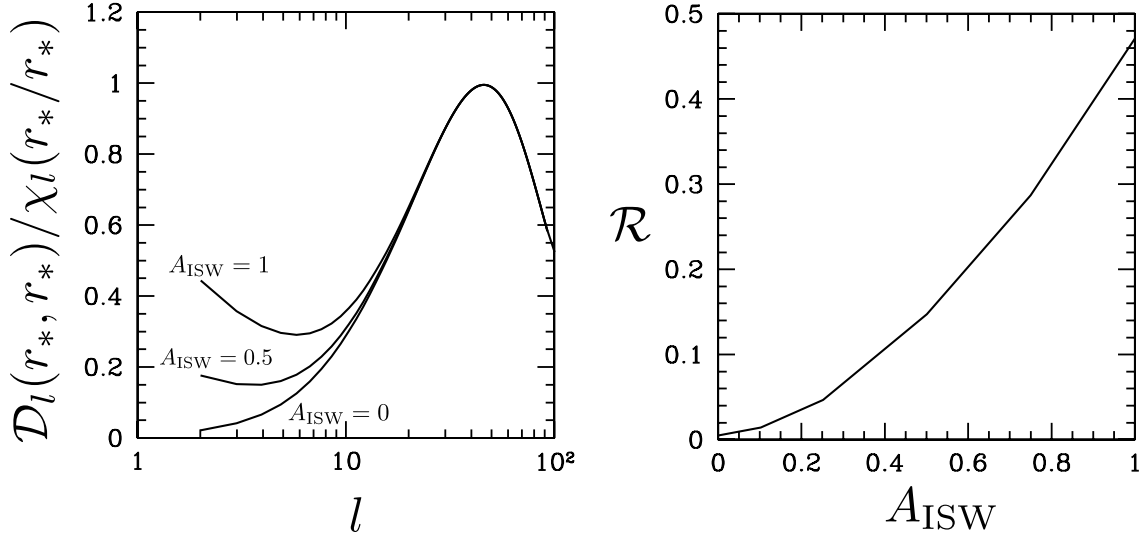
FIG. 4: *Left*: The realization-dependent normalization reduces the $f_{\rm NL}$-dependent variance to varying extents depending on the level of the late-time ISW effect. From top to bottom we plot the function $D_l(r_*, r_*)/\chi_l(r_*, r_*)$ with varying levels of late-time ISW: 1, 0.5, 0. The overall reduction of the $f_{\rm NL}$-dependent variance roughly corresponds to the value of $\mathcal{D}_2(r_*, r_*)/\chi_2(r_*, r_*)$. *Right*: The fractional improvement in the $f_{\rm NL}$-dependent variance, $\mathcal{R}$, as a function of the amplitude of the late-time ISW. When the late-time ISW is completely removed (i.e., $A_{\rm ISW} = 0$) the $f_{\rm NL}$-dependent variance is nearly completely removed; for an unmodified ISW level the $f_{\rm NL}$-dependent variance is reduced by a factor of ~ 0.5.

it is zero the late-time ISW effect is absent. In Fig. 4 we show $\mathcal{D}_l(r_*, r_*)/\chi_l(r_*, r_*)$ as a function of multipole, l, with varying late-time ISW amplitude, $A_{\rm ISW}$. On the right hand side of the figure, we can see $\mathcal{R} \simeq \mathcal{D}_2(r_*, r_*)/\chi_2(r_*, r_*)$ falls off sharply as $A_{\rm ISW}$ decreases. Qualitatively this result can be understood by noting that with both the Sachs-Wolfe and late-time ISW effects contributing to the large-scale anisotropies our estimator $\widehat{a}_{lm}^{\rm NL}$ becomes less correlated with the actual $a_{lm}^{\rm NL}$ leading to a larger value for the fractional reduction $\mathcal{R}$.

In order to better understand how the late-time ISW effect limits our ability to remove the $f_{\rm NL}$-dependent variance quantitatively we can approximate the anisotropies on large-scales by [51]

$$a_{lm} \simeq -\frac{1}{3}\Phi_{lm}(r_*) + A_{\rm ISW}\Phi_{lm}(r_{\rm ISW}), \tag{71}$$

where $A_{\rm ISW}$ controls the contribution of the ISW effect to the temperature multipoles, and $r_{\rm ISW}$ is the conformal distance from the observer to the peak of the ISW visibility function. The ISW contribution extends from ~ 500 Mpc to 4000 Mpc (where the present time is 0 Mpc). Eq. (71) explicitly shows how the contribution of the late-time ISW effect to the observed multipole moments a_{lm} limits our ability to reconstruct $a_{lm}^{\rm NL}$. Since the correlation between the Sachs-Wolfe and ISW terms is at most 10% we neglect it and we have

$$\mathcal{R} \simeq \frac{\mathcal{D}_2(r_*, r_*)}{\chi_2(r_*, r_*)} \simeq 1 - \frac{C_l^{\rm SW}}{C_l^{\rm SW} + A_{\rm ISW}^2 C_2^{\rm ISW}}, \tag{72}$$

where $C_l^{\rm SW}$ and $C_l^{\rm ISW}$ are the Sachs-Wolfe and the late-time ISW contribution to the power spectrum, respectively. As shown on the right-hand panel in Fig. 4, in a ΛCDM universe with all of the late-time ISW effect included, the $f_{\rm NL}$-dependent variance can only be reduced by a factor of 0.5 using the realization-dependent normalization defined in Eqs. (39) and (43). When the late-time ISW contribution is completely removed, the $f_{\rm NL}$-dependent variance is reduced by a factor of 20 ($\mathcal{R} = 0.05$). This raises the question of whether or not probes of large-scale structure closer to the present epoch can be used to 'remove' the late-time ISW contribution to the CMB anisotropies and further reduce the $f_{\rm NL}$-dependent variance. Next we discuss how we may use such a tracer of the ISW effect to further reduce the $f_{\rm NL}$-dependent variance of the RNE.

VI. ISW SUBTRACTION WITH FOREGROUND TRACERS

We have identified the late-time ISW effect as the cause of the residual variance scaling as $f_{\rm NL}^2$. If the late-time ISW component of a temperature map could be estimated, and then cleaned out from the data, we expect that the

corresponding generalization of the RNE would then be nearly free of the f_{NL}-dependent variance.

Since the bulk of the late-time ISW effect in a ΛCDM cosmology comes from redshift $z \sim 1$, a measurement of the large-scale gravitational potential or density field around this redshift should yield information that we can use to clean our map of the ISW effect. The use of an ISW-cleaned map to improve the sensitivity of the CMB to primordial non-Gaussianity was first proposed in Ref. [48]. In that work, only the zeroth order variance $\langle (\Delta \mathcal{B}_0)^2 \rangle$ was computed, and as a result the improvement in the S/N obtained by using ISW-cleaned maps was marginal. In contrast, here we will see that using a large-scale structure tracer has the potential to reduce the f_{NL}-dependent variance by up to $\sim$ 90%.

For an arbitrary large-scale tracer t_{lm}, the ISW-cleaned CMB temperature anisotropy is

$$a_{lm}^{\mathrm{c}} = a_{lm} - \frac{\langle a_{lm}^{\mathrm{ISW}} t_{lm}^* \rangle}{\langle t_{lm}\, t_{lm}^* \rangle} t_{lm}, \tag{73}$$

where a_{lm}^{ISW} is the portion of the total multipoles, a_{lm}, due to the late-time ISW effect, t_{lm} is the multipole associated with the large-scale tracer field, a superscript c indicates a quantity that has been 'cleaned' of the late-time ISW effect, and a superscript t indicates a quantity associated with the tracer field. We note that if we use the lensing potential for CMB weak lensing as our tracer field then t_{lm} is obtained from some higher-order cumulant of the data. This then implies that the tracer's power spectrum must also take into account an additional noise term, as we discuss in more detail below.

The RNE takes the same form as in Eq. (39) but written in terms of ISW-cleaned quantities:

$$C_l^{\mathrm{c}} \;\equiv\; C_l - 2\frac{C_l^{\mathrm{T,t}} C_l^{\mathrm{ISW,t}}}{C_l^{\mathrm{tt}}} + \frac{\left(C_l^{\mathrm{ISW,t}} \right)^2}{C_l^{\mathrm{tt}}}, \tag{74}$$

$$\alpha_l^{\mathrm{c}}(r) \;\equiv\; \alpha_l(r) - \alpha_l^{\mathrm{t}}(r), \tag{75}$$

$$\beta_l^{\mathrm{c}}(r) \;\equiv\; \beta_l(r) - \beta_l^{\mathrm{t}}(r), \tag{76}$$

where $\alpha_l^{\mathrm{t}}(r)$ and $\beta_l^{\mathrm{t}}(r)$ are defined as in Eqs. (11) and (12) but with in terms of a transfer function corresponding to the power spectrum of the tracer field. In order to evaluate the fractional reduction of the f_{nl}-dependent variance with an ISW cleaned CMB map, we evaluate Eq. (61) making the identification $\alpha_l(r) \to \alpha_l^{\mathrm{c}}(r)$, $\beta_l(r) \to \beta_l^{\mathrm{c}}(r)$, and $C_l \to C_l^{\mathrm{c}}$. Now that we have expressions for the fractional reduction of the f_{NL}-dependent variance of an estimator $\widehat{f_{\mathrm{NL}}}^{\mathrm{c}}$ built from ISW-cleaned maps, we consider two specific tracers of foreground structure: weak lensing of the CMB and galaxy surveys.

A. Reconstructed lensing potential

Weak lensing deflects the trajectories of CMB photons, remapping the temperature and polarization fields. The projected potential $\Psi(\hat{n})$ determines the trajectories of lensed CMB photons. A reconstruction of the projected potential field from measurements of the CMB would allow a separate probe of large-scale structure up to redshifts of $\sim$ a few.

Referring to the solid curves in the left-hand panel of Fig. 5 we show the correlation coefficient between the lensing potential and the late-time ISW effect, as well as the correlation coefficient between the deflection and full temperature fields. We see that the multipole moments of the deflection field are very strongly correlated with the late-time ISW effect. The fractional reduction of variance obtained with the realization-dependent normalization for the perfectly (via lensing) cleaned maps is controlled by the ratio $\mathcal{D}_l^{\Psi}(r_*, r_*)/\chi_l(r_*, r_*)$ is shown in the right-hand panel of Fig. 5. We find that the realization-dependent normalization removes $\sim$ 90% of the f_{NL}^2 variance if the map is 'cleaned' of the late-time ISW effect using a perfect reconstruction of the CMB deflection potential.

Of course, even with a idealized noiseless CMB map, the deflection potential can not be perfectly reconstructed. The lensing estimator relies on off-diagonal correlations of temperature multipoles, and in any given realization of the power spectrum, there will be some overlap between the lensing estimator and chance correlations that arise from cosmic variance. This leads to the reconstruction noise variance N_l [48, 52].

On the large scales relevant for ISW subtraction, the bulk of the S/N for lensing construction comes from very high l, and so instrument noise can be neglected in our estimates. We make the replacement $C_l^{\Psi\Psi} \to C_l^{\Psi\Psi} + N_l$ above to assess how much a realistic reconstruction of Ψ (and subsequent cleaning of the CMB temperature map) could improve the performance of the RNE. We assume a reference experiment with temperature and polarization noise $\Delta T = \Delta P/\sqrt{2} = 1\ \mu$K arcmin, and an angular resolution of $\sigma = 4'$, as discussed in detail in Ref. [53]. We note that measurements using Planck will not be sensitive enough to realistically use the reconstructed lensing potential in
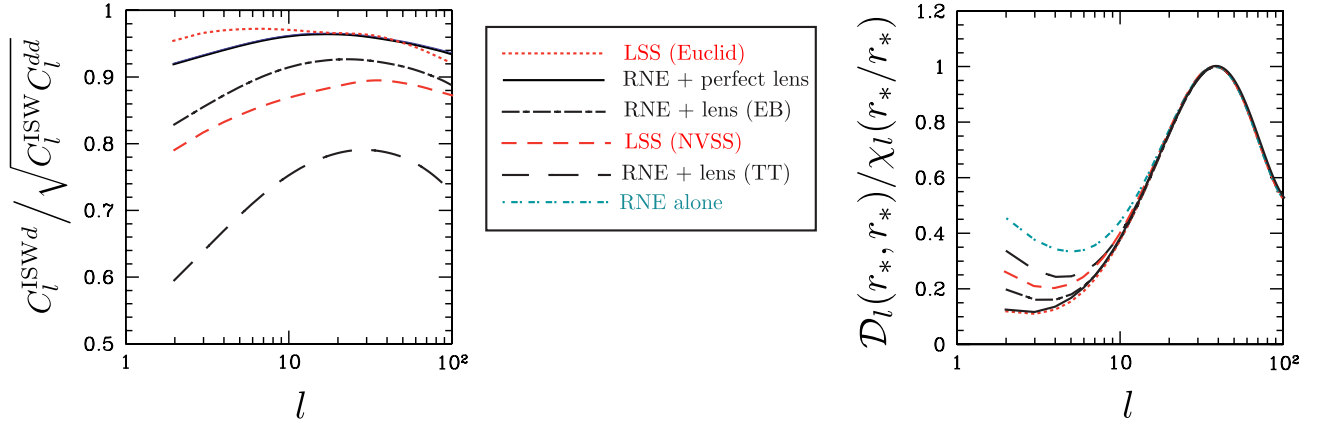
FIG. 5: *Left*: The three black curves show the correlation between the lensing potential and the late-time ISW effect in the CMB temperature: the solid curve shows the correlation without taking into account lensing reconstruction noise; the dot-dashed curve shows the correlation between the late-time ISW effect and the reconstructed lensing potential using EB correlations, including reconstruction noise; analogously the long-dashed curve shows the same correlation, but using TT correlations to reconstruct the lensing potential. The two red curves show the correlation between the forecasted Euclid/WFIRST (dotted) or NVSS (medium dashed) density fields, and the late-time ISW effect. *Right*: The fractional reduction in the quantity $\mathcal{D}_l$ for the uncleaned CMB maps (cyan, medium-short dashed) and cleaned maps with line-types corresponding to the curves in the left panel.

order to remove the late-time ISW effect. From the right-hand panel of Fig. 5, we see that if the temperature field is used as to reconstruct the deflection field, the RNE can remove $\sim 70\%$ of the variance proportional to f_{NL}^2, whereas if the polarization field (through the EB correlation) is used, the RNE can remove $\sim 80\%$ of the excess variance[2].

B. Galaxy survey

Alternatively, a galaxy survey with peak redshift near $z \sim 1$ also probes the potential field at the epochs when the late-time ISW effect is imprinted on the CMB. For a galaxy survey with selection function $w(z)$, the transfer-function is given by

$$S_l^g(k) \equiv \int_0^{r_*} dr b_k(r) S_k^m(r) w[z(r)] j_l(kr) \tag{77}$$

where $b_k(r)$ is the bias, $z(r)$ is the redshift as a function of conformal distance, and $S_k^m(r)$ is the time and scale-dependent function which maps the primordial potential to the evolved matter density. We have taken [41, 42]

$$w(z) = CH(z) \left(\frac{z^m}{z_0^{m+1}} \right) \exp\left[-\left(\frac{z}{z_0} \right)^\beta \right], \tag{78}$$

where for the NVSS survey we take $z_0 = 0.79$, $\beta = 1$, and $m = 1.18$ as in Ref. [54]. On the other hand, for a futuristic space-based all-sky dedicated dark energy, such as that described in the Joint Dark Energy Mission (WFIRST) [55] and Euclid mission concept [56], or the Large Synoptic Survey Telescope [55] (LSST) we take $z_0 = 0.5$, $\beta = 1$, $m = 2$ [55]. The presence of the Hubble parameter converts the selection function from galaxies per unit redshift, to galaxies per unit conformal time. We do not need to specify the normalization constant C, as it divides out of all quantities of interest.

---

[2] Of course in a more complete analysis, we could include other correlation functions (EE, TE, EE, and BB) in the reconstruction of the deflection field. We choose to focus on TT (since TT-based detection of CMB weak lensing have already been made) and EB (because it vanishes in the null hypothesis of no lensing, and thus provides a strong probe of the deflection field). These other correlation functions could themselves be used to look for non-Gaussianity– indeed, Refs. [23] and [43] introduce the MVNH that includes all of them. In this work, we restrict ourself to the simple case discussed already and leave a more complete analysis for future work.

We approximate the bias as constant in scale as well as in redshift [41, 42] so that it factors out of the computation completely[3]. Referring to the dashed-lines in Fig. 5, we show the correlation coefficient between large-scale structure (for both NVSS and Euclid/WFIRST parameters) and the late-time ISW effect, as well as the correlation coefficient between large-scale structure and the full temperature field. We see that the multipole moments of the large-scale density field are very strongly correlated with the late-time ISW effect.

The fractional reduction in the variance obtained with the RNE for the cleaned maps is shown in the right-hand panel of Fig. 5. We find that the realization-dependent normalization removes $\sim 75\%$ of the f_{NL}^2 variance if the ISW effect is removed by cross-correlating a CMB map with a large-scale structure survey for the NVSS survey parameters, and $\sim 90\%$ of the f_{NL}^2 variance if WFIRST survey parameters are assumed. In Fig. 6, we see how the S/N for f_{NL} is improved if prior to the application of the realization-dependent normalization, the late-time ISW effect is removed from maps using two different secondary tracers. This figure was produced assuming the variance of the standard MVNH estimator given in Eq. (18) and the reduction in the f_{NL}-dependent variance, $\mathcal{R}$, using various large-scale structure tracers to 'clean' the CMB maps of the late-time ISW effect,

$$\frac{S}{N} = \frac{f_{\mathrm{NL}}}{\sqrt{\frac{1}{72Af_{\mathrm{sky}}l_{\max}^2 \ln(l_{\max})} + \mathcal{R}\frac{f_{\mathrm{NL}}^2}{2\ln^3(l_{\max})}}}. \tag{79}$$

Although the scaling with $l_{\max}$ and f_{NL} given in the above equation was calculated in the flat-sky Sachs-Wolfe limit [31, 33, 47], it has been shown to reproduce the $l_{\max}$ scaling calculated on the full sky and with the full transfer-function [23, 32]. However, we note that the exact numerical factors may be different for the full sky and full transfer-function case. It is thus possible that the f_{NL}-dependent variance may be even more important at lower f_{NL} than is indicated by this expression. As future data is used to determine the level of non-Gaussianity in the CMB, the statistics of the standard null-hypothesis estimator should be checked, especially if the data starts to indicate that $f_{\mathrm{NL}} \neq 0$.

We see from this figure that if primordial non-Gaussianity is of local type and $f_{\mathrm{NL}} \gtrsim 5$, a significant improvement in the S/N for f_{NL} may be obtained by using cleaned maps. Of course in a real galaxy survey, there are additional complications due to incomplete sky coverage, photometric redshift errors, and shot-noise due to a finite number of galaxies in the survey volume. Here we neglect these important real-world effects to highlight the fact that a measurement of the ISW effect can help reduce the f_{NL}^2 variance in the RNE. An increase in the S/N in an estimate for f_{NL} would not only lead to a more precise determination of the level of non-Gaussianity in the CMB, but would also lead to a more precise estimation of additional parameters (such as a possible scale-dependence) associated with primodial non-Gaussianity.

VII. CONCLUSIONS

We have investigated the origin of the f_{NL}-dependent variance when applying the standard MVNH f_{NL} estimator to CMB maps with appreciable non-Gaussianity. We found that this variance is due to terms that appear in the estimator which do not contribute to the signal but which do contribute to the noise.

Previous work in Ref. [34] has shown that a Bayesian analysis has the potential to provide an estimate of f_{NL} from the CMB which does not show an f_{NL}-dependent increase in the variance when applied to maps with appreciable non-Gaussianity. That approach, however, is computationally expensive and quite inefficient, taking 150,000 CPU hours to compute the estimator on a simulated non-Gaussian CMB map. It is therefore desirable to find a computationally simple and efficient method to estimate f_{NL} from the CMB which remains optimal even when applied to maps with appreciable non-Gaussianity.

We have found that a new realization-dependent estimator (RNE) can be constructed which is computationally efficient (utilizing the scaling properties of fast-Fourier transforms) and reduces the f_{NL}-dependent variance by a factor of ~ 2. Previous studies have shown this same realization-dependent normalization can completely remove the f_{NL}-dependent variance in the Sachs-Wolfe limit. When the full transfer-function is used, however, this limit is a poor approximation to the CMB power-spectrum, especially at large-scales where the late-time ISW effect (due to late-time acceleration) contributes about half of the power. We have artificially reduced the level of the late-time ISW and found that when it is completely removed our new estimator has negligible f_{NL}-dependent variance even with the full transfer function. This implies that by using a tracer which effectively removes the late-time ISW contribution to the CMB map we can use the RNE to further reduce the f_{NL}-dependence.

---

[3] In reality, local-type non-Gaussianity would also induce scale-dependent bias [57], and thus corrections to the variance of the cleaned map. This effect, however, would be higher order in f_{NL}, so we may neglect it without loss of generality.
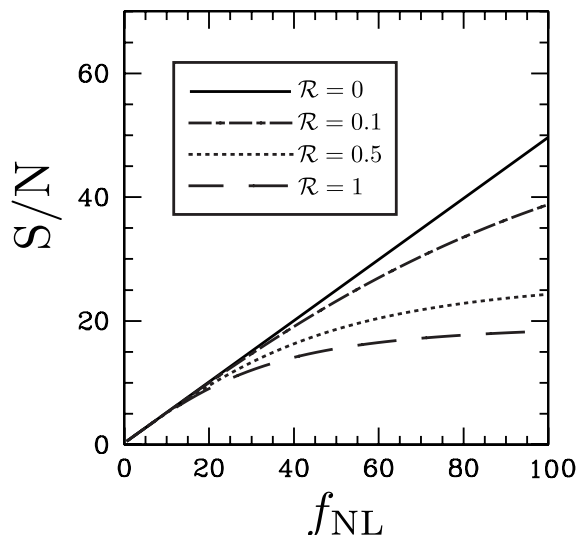
FIG. 6: The signal to noise (S/N) in an experiment with $l_{\max} = 2500$ and $f_{\mathrm{sky}} = 0.8$ (corresponding to Planck) as a function of f_{NL} under the flat-sky and Sachs-Wolfe approximations. The top (solid) curve shows the S/N for the standard MVNH estimator without taking into account the f_{NL}-dependent variance. The curve second from the top (dot-dashed) shows the S/N for the RNE after a tracer field has been used to remove most of the late-time ISW effect, leading to a reduction of the f_{NL}-dependent variance by a factor of $\mathcal{R} = 0.1$. The curve third from the top (dotted) shows the S/N for the RNE using only CMB data with $\mathcal{R} = 0.5$. The bottom curve (long-dashed) shows the S/N for the standard MVNH estimator with the (full) f_{NL}-dependent variance with $\mathcal{R} = 1$. As this figure shows, the cleaned maps can increase the S/N by several standard deviations.

We considered two tracers of large-scale structure: the deflection field (which generates lensing in the CMB) and a large-scale structure survey (considering survey parameters comparable to those of NVSS and Euclid/WFIRST) with a mean redshift of $z \sim 1$. Both tracers are highly correlated with the late-time ISW. We find that by using the deflection field as a tracer, we can reduce the f_{NL}-dependent variance by a factor of ~ 0.1 using a futuristic CMB experiment (the reconstruction of the deflection potential from Planck is not accurate enough to be useful). If the large-scale structure measured by NVSS is used as a tracer, the variance could be reduced by a factor of ~ 0.25, while a next-generation mission like Euclid/WFIRST could reduce the variance by a factor of ~ 0.1.

We show the improvement in the S/N for the estimation of f_{NL} using the RNE assuming a satellite experiment like Planck ($l_{\max} = 2500$, $f_{\mathrm{sky}} = 0.8$) in Fig. 6. If the data indicate that $f_{\mathrm{NL}} \neq 0$, then the estimator discussed here can be used to increase the S/N of the detection, thus shedding new light on primordial non-Gaussianity.

Appendix A: Flat-sky approximation

To speed computation throughout the paper (by allowing the use of FFTs along both dimensions of the sky), we use the flat-sky approximation as described in Refs. [33, 58], while providing whole-sky expressions for use on real maps. In terms of the fractional temperature perturbation $T(\hat{n})$ at position $\hat{n}$, the temperature power spectrum C_l is

given by

$$\left\langle a_{\vec{l_1}} a_{\vec{l_2}} \right\rangle = \Omega \delta_{\vec{l_1}+\vec{l_2},0} C_l, \tag{A1}$$

$$a_{\vec{l}} = \int d^2\vec{\theta}\, e^{-i\vec{l}\cdot\vec{\theta}} T(\vec{\theta}) \simeq \frac{\Omega}{N_{\text{pix}}} \sum_{\vec{\theta}} e^{-i\vec{l}\cdot\vec{\theta}} T(\vec{\theta}), \tag{A2}$$

where $\Omega = 4\pi f_{\text{sky}}$ is the survey area (in steradians), and $\delta_{\vec{l_1}+\vec{l_2},0}$ is a Kronecker delta that sets $\vec{l_1} = -\vec{l_2}$ and we use the convention $\delta^2(\vec{l}) = \Omega \delta_{\vec{l}}$ to convert between Kronecker and Dirac-δ functions. We convert between discrete sums and integrals using the correspondence $\Sigma_{\vec{l}} \leftrightarrow \Omega \int d^2\vec{l}/(2\pi)^2$ [58]. We use the convention

$$P(k) = Ak^{n_{\text{s}}-4}, \tag{A3}$$

and assume a scale-invariant power spectrum ($n_s = 1$) for the duration of this paper. For a scale-invariant primordial power spectrum in the Sachs-Wolfe approximation, the angular power spectrum for $T(\vec{\theta})$ is [23]:

$$C_l = \frac{A}{9\pi l(l+1)}, \tag{A4}$$

where we take the amplitude $A = 2\pi^2\Delta_\Phi^2 \simeq 2.43 \times 10^{-9} \times 2\pi^2 \simeq 4.7 \times 10^{-8}$ [1]. In the flat-sky limit, the reduced bispectrum $b_{l_1 l_2 l_3}$ is given by

$$\left\langle a_{\vec{l_1}} a_{\vec{l_2}} a_{\vec{l_3}} \right\rangle = \Omega \delta_{\vec{l_1}+\vec{l_2}+\vec{l_3},0} b_{l_1 l_2 l_3}. \tag{A5}$$

The expression for $b_{l_1 l_2 l_3}$ in Eq. (10) itself [in terms of $\alpha_l(r)$, $\beta_l(r)$] is unchanged. The Kronecker delta insures that the bispectrum is defined only for $\vec{l_1} + \vec{l_2} + \vec{l_3} = 0$; i.e., only for triangles in Fourier space. Statistical isotropy then dictates that the bispectrum depends only on the magnitudes l_1, l_2, l_3 of the three sides of this Fourier triangle. To derive various expressions in the text, we will use the flat-sky equivalent of the Wigner-3J coefficient [52]:

$$\sqrt{\frac{(2l_1+1)(2l_2+1)(2l_3+1)}{4\pi}} \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \rightarrow \delta_{\vec{l_1}+\vec{l_2}+\vec{l_3},0}\Omega. \tag{A6}$$

In the flat-sky limit, the MVNH is given (applying the same arguments used to derive it in the whole-sky case) by [23, 58]

$$\widehat{f_{\text{NL}}} \equiv \sigma_0^2 \sum_{\vec{l_1}+\vec{l_2}+\vec{l_3}=0} \frac{a_{\vec{l_1}} a_{\vec{l_2}} a_{\vec{l_3}} b_{l_1 l_2 l_3}}{6\Omega^2 C_{l_1} C_{l_2} C_{l_3}}, \tag{A7}$$

and it has inverse variance,

$$\sigma_0^{-2} = \sum_{\vec{l_1}+\vec{l_2}+\vec{l_3}=0} \frac{[b_{l_1 l_2 l_3}]^2}{6\Omega C_{l_1} C_{l_2} C_{l_3}}. \tag{A8}$$

Appendix B: Details of the calculation of the variance

We will first compute the variance of $\mathcal{B}_1$. To do so we will concentrate on the part of the variance

$$\int r^2 dr (r')^2 dr' \alpha_{l_1}(r) \alpha_{m_1}(r') \left\langle a_{\vec{l_2}} a_{\vec{l_3}} \phi_{\vec{k}}(r) \phi_{\vec{k'}}(r') | a_{\vec{m_2}}^* a_{\vec{m_3}}^* \phi_{\vec{p}}^*(r') \phi_{\vec{p'}}^*(r') \right\rangle$$

$$\times \quad \delta_{\vec{l_1}+\vec{l_2}+\vec{l_3},0} \delta_{\vec{k}+\vec{k'},\vec{l_1}} \delta_{\vec{m_1}+\vec{m_2}+\vec{m_3},0} \delta_{\vec{p}+\vec{p'},\vec{m_1}}. \tag{B1}$$

We will first concentrate on identifying the various types of terms that will arise from the expectation value. First note that the Kronecker deltas require that the only 'internal' contractions can be between the $\vec{l}$ or $\vec{m}$ terms with the

$\vec{k}$s or $\vec{p}$s. Therefore the only contractions on the 'off-diagonal' [in which all $a_{\vec{l}}$s are contracted with $\phi_{l'}(r)$] are:

$$A_1 = \left\langle a_{\vec{l}_2}\phi_{\vec{k}}(r)\right\rangle\left\langle a^*_{\vec{m}_2}\phi^*_{\vec{p}}(r')\right\rangle\left\langle a_{\vec{l}_3}a^*_{\vec{m}_3}\right\rangle\left\langle \phi_{\vec{k}'}(r)\phi_{\vec{p}'}(r')^*\right\rangle, \tag{B2}$$

$$= \delta_{\vec{l}_2,-\vec{k}}\beta_{l_2}(r)\delta_{\vec{m}_2,-\vec{p}}\beta_{m_2}(r')C_{l_3}\delta_{\vec{l}_3,\vec{m}_3}\chi_{k'}(r,r')\delta_{\vec{k}',\vec{p}'}, \tag{B3}$$

$$A_2 = \left\langle a_{\vec{l}_2}\phi_{\vec{k}}(r)\right\rangle\left\langle a^*_{\vec{m}_2}\phi^*_{\vec{p}}(r')\right\rangle\left\langle a_{\vec{l}_3}\phi^*_{\vec{p}}(r')\right\rangle\left\langle \phi_{\vec{k}'}(r)a^*_{\vec{m}_3}\right\rangle, \tag{B4}$$

$$= \delta_{\vec{l}_2,-\vec{k}}\beta_{l_2}(r)\delta_{\vec{m}_2,-\vec{p}}\beta_{m_2}(r')\delta_{\vec{l}_3,\vec{p}'}\beta_{l_3}(r')\delta_{\vec{m}_3,\vec{k}'}\beta_{m_3}(r). \tag{B5}$$

There are 4!=24 'diagonal' contractions; however, not all are unique since the sum is symmetric in $(\vec{l}_1,\vec{l}_2)$, $(\vec{k},\vec{k}')$, $(\vec{m}_1,\vec{m}_2)$, $(\vec{p},\vec{p}')$. Representing these pairs by numbered boxes in Fig. 7 we show the five unique combinations that will make up the variance. The other three unique terms are:
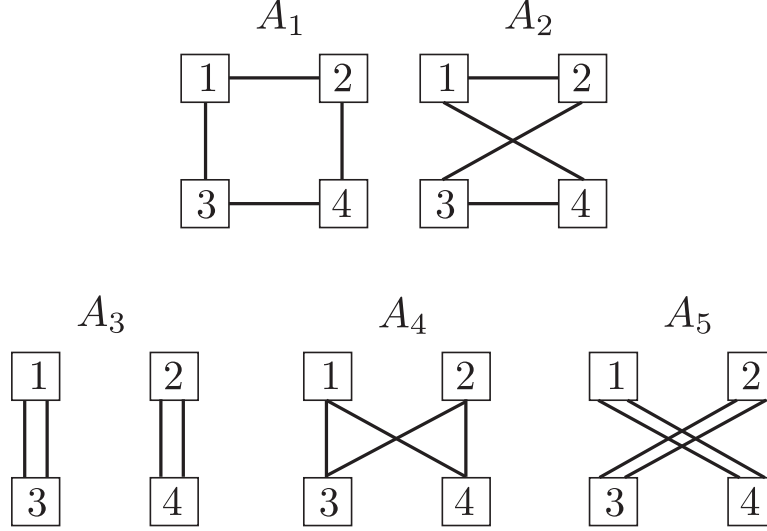

FIG. 7: A graphical representation of the 5 combinations of $1 = (\vec{l}_2,\vec{l}_3)$, $2 = (\vec{k},\vec{k}')$, $3 = (\vec{m}_2,\vec{m}_3)$, $4 = (\vec{p},\vec{p}')$ which make up the variance of $\mathcal{B}_1$.

$$A_3 = \left\langle a_{\vec{l}_2}a^*_{\vec{m}_2}\right\rangle\left\langle a_{\vec{l}_3}a^*_{\vec{m}_3}\right\rangle\left\langle \phi_{\vec{k}}(r)\phi^*_{\vec{p}}(r')\right\rangle\left\langle \phi_{\vec{k}'}(r)\phi^*_{\vec{p}'}(r')\right\rangle \tag{B6}$$

$$= C_{l_2}\delta_{\vec{l}_2,\vec{m}_2}C_{l_3}\delta_{\vec{l}_3,\vec{m}_3}\chi_k(r,r')\delta_{\vec{k},\vec{p}}\chi_{k'}(r,r')\delta_{\vec{k}',\vec{p}'}. \tag{B7}$$

$$A_4 = \left\langle a_{\vec{l}_2}a^*_{\vec{m}_2}\right\rangle\left\langle a_{\vec{l}_3}\phi^*_{\vec{p}}(r')\right\rangle\left\langle \phi_{\vec{k}}(r)a^*_{\vec{m}_3}\right\rangle\left\langle \phi_{\vec{k}'}(r)\phi^*_{\vec{p}'}(r)\right\rangle \tag{B8}$$

$$= C_{l_2}\delta_{\vec{l}_2,\vec{m}_2}\beta_{l_3}(r')\delta_{\vec{l}_3,\vec{p}}\beta_{m_3}(r)\delta_{\vec{k},\vec{m}_3}\chi_{k'}(r,r')\delta_{\vec{k}',\vec{p}'}. \tag{B9}$$

$$A_5 = \left\langle a_{\vec{l}_2}\phi^*_{\vec{p}}(r')\right\rangle\left\langle a_{\vec{l}_3}\phi^*_{\vec{p}'}(r)\right\rangle\left\langle \phi_{\vec{k}}(r)a^*_{\vec{m}_2}\right\rangle\left\langle \phi_{\vec{k}'}(r')a^*_{\vec{m}_3}\right\rangle \tag{B10}$$

$$= \beta_{l_2}(r')\delta_{\vec{l}_2,\vec{p}}\beta_{l_3}(r)\delta_{\vec{l}_3,\vec{p}'}\beta_{m_2}(r)\delta_{\vec{k},\vec{m}_2}\beta_{m_3}(r')\delta_{\vec{m}_3,\vec{k}'}. \tag{B11}$$

Therefore, in the end we have five unique combinations:

$$A_1 = C_{l_3}\delta_{\vec{l}_3,\vec{m}_3}\beta_{l_2}(r)\delta_{\vec{l}_2,-\vec{k}}\beta_{m_2}(r')\delta_{\vec{m}_2,-\vec{p}}\chi_{k'}(r,r')\delta_{\vec{k}',\vec{p}'}, \tag{B12}$$

$$A_2 = \beta_{l_2}(r)\delta_{\vec{l}_2,-\vec{k}}\beta_{m_3}(r)\delta_{\vec{m}_3,\vec{k}'}\beta_{l_3}(r')\delta_{\vec{l}_3,\vec{p}'}\beta_{m_2}(r')\delta_{\vec{m}_2,-\vec{p}}, \tag{B13}$$

$$A_3 = C_{l_2}\delta_{\vec{l}_2,\vec{m}_2}C_{l_3}\delta_{\vec{l}_3,\vec{m}_3}\chi_k(r,r')\delta_{\vec{k},\vec{p}}\chi_{k'}(r,r')\delta_{\vec{k}',\vec{p}'}, \tag{B14}$$

$$A_4 = C_{l_2}\delta_{\vec{l}_2,\vec{m}_2}\beta_{m_3}(r)\delta_{\vec{m}_3,\vec{k}}\beta_{l_3}(r')\delta_{\vec{l}_3,\vec{p}}\chi_{k'}(r,r')\delta_{\vec{k}',\vec{p}'}, \tag{B15}$$

$$A_5 = \beta_{l_3}(r)\delta_{\vec{l}_3,\vec{p}'}\beta_{m_2}(r)\delta_{\vec{m}_2,\vec{k}}\beta_{l_2}(r')\delta_{\vec{l}_2,\vec{p}}\beta_{m_3}(r')\delta_{\vec{m}_3,\vec{k}'}. \tag{B16}$$

The last term, A_5, has the Kronecker deltas $\delta_{\vec{l}_3,\vec{p}'}\delta_{\vec{l}_2,\vec{p}}$ which implies the full term will have the Kronecker delta $\delta_{\vec{l}_1+\vec{l}_2+\vec{l}_3,0}\delta_{\vec{p}+\vec{p}',\vec{l}_1}\delta_{\vec{l}_3,\vec{p}'}\delta_{\vec{l}_2,\vec{p}}$ so that when summing over $\vec{p}$ and $\vec{p}'$ we will have $\vec{l}_2=\vec{p}$ and $\vec{l}_3=\vec{p}'$ so the final term will be zero since $|\vec{l}_1|\geq 2$. Therefore $A_5=0$ and we are left with four unique terms, which agrees with the appendix in Ref. [33]. The same approach can be taken with the variance of $\widehat{\mathcal{B}}_1$

$$\left\langle (\Delta\widehat{\mathcal{B}}_1)^2 \right\rangle \sim \frac{\beta_k(r)\beta_{k'}(r)}{C_k C_{k'}}\frac{\beta_p(r')\beta_{p'}(r')}{C_p C_{p'}}\left\langle a_{\vec{l}_2}a_{\vec{l}_3}a_{\vec{k}}a_{\vec{k}'}|a^*_{\vec{m}_2}a^*_{\vec{m}_3}a^*_{\vec{p}}a^*_{\vec{p}'}\right\rangle, \tag{B17}$$

and the covariance between $\widehat{\mathcal{B}}_1$ and $\mathcal{B}_1$,

$$\left\langle \Delta\widehat{\mathcal{B}}_1\Delta\mathcal{B}_1 \right\rangle \sim \frac{\beta_k(r)\beta_{k'}(r)}{C_k C_{k'}}\left\langle a_{\vec{l}_2}a_{\vec{l}_3}a_{\vec{k}}a_{\vec{k}'}|a^*_{\vec{m}_2}a^*_{\vec{m}_3}\phi^*_{\vec{p}}(r')\phi^*_{\vec{p}'}(r')\right\rangle. \tag{B18}$$

Computing these terms shows that certain contractions dominate the sum so that [33]

$$\begin{aligned}
\langle\Delta(\mathcal{B}_1)\Delta(\mathcal{B}_1)\rangle &= 8\sum_{\{\vec{l}\},\{\vec{k}\}}\frac{b_{l_1l_2l_3}b_{k_1k_2k_3}}{C_{l_1}C_{l_2}C_{k_1}C_{k_2}C_{k_3}}\delta_{\vec{l}_3,\vec{k}_3} \\
&\quad \times \int r^2 dr(r')^2 dr'\alpha_{l_1}(r)\alpha_{k_1}(r')\beta_{k_2}(r')\beta_{l_2}(r)\chi_{l_3}(r,r'),
\end{aligned} \tag{B19}$$

$$\begin{aligned}
\left\langle\Delta(\widehat{\mathcal{B}_1})\Delta(\widehat{\mathcal{B}_1})\right\rangle &= 8\sum_{\{\vec{l}\},\{\vec{k}\}}\frac{b_{l_1l_2l_3}b_{k_1k_2k_3}}{C_{l_1}C_{l_2}C_{k_1}C_{k_2}C_{k_3}}\delta_{\vec{l}_3,\vec{k}_3} \\
&\quad \times \int r^2 dr(r')^2 dr'\alpha_{l_1}(r)\alpha_{k_1}(r')\beta_{k_2}(r')\beta_{l_2}(r)\frac{\beta_{l_3}(r)\beta_{l_3}(r')}{C_{l_3}},
\end{aligned} \tag{B20}$$

$$\left\langle\Delta(\mathcal{B}_1)\Delta(\widehat{\mathcal{B}_1})\right\rangle = \langle\Delta(\mathcal{B}_1)\Delta(\mathcal{B}_1)\rangle. \tag{B21}$$

Appendix C: Fast algorithm to compute RNE

As noted in Refs. [26, 28], due to the inefficiency of harmonic transforms, the MVNH estimator is expensive to evaluate, requiring the computation of $\sim l_{\max}^5$ terms. However, a more efficient computational algorithm can be used once the MVNH estimator is written in terms real-space quantities- once this is done, the azimuthal part of the harmonic transform can be computed using computationally efficient fast-Fourier transforms (FFTs). As is noted in Ref. [26] the MVNH estimator can be rewritten as

$$\widehat{f_{\mathrm{NL}}} = \sigma_0^{-2}\int r^2 dr\int d^2\hat{n}B^2(\hat{n},r)A(\hat{n},r), \tag{C1}$$

$$A(\hat{n},r) = \sum_{lm}\frac{\alpha_l(r)Y_{lm}(\hat{n})a_{lm}}{C_l}, \tag{C2}$$

$$B(\hat{n},r) = \sum_{lm}\frac{\beta_l(r)Y_{lm}(\hat{n})a_{lm}}{C_l}. \tag{C3}$$

At each location along the line of sight, the resulting estimator only requires the computation of $\sim l_{\max}^3$ terms. In addition to this, the filter functions $\alpha_l(r)$ and $\beta_l(r)$ are sufficiently smooth so that they must only be evaluated for $\mathcal{O}(100)$ grid points. In Sec. IV, we generalized the realization-dependent normalization of Ref. [31] to treat the full sky and include the radiation transfer function. Here we show how we may rewrite this estimator in order to utilize FFTs to speed up their computation.

The estimator is given by

$$\widehat{\mathcal{B}}_1 = \sigma_0^{-2}\sum_{l\leq l_1\leq l_2, l_a l_b}\sum_{m_1 m_2 m_a m_b m}\frac{a_{l_1 m_1}a_{l_2 m_2}a_{l_a m_a}a_{l_b m_b}}{2C_l C_{l_1}C_{l_2}C_{l_a}C_{l_b}}B_{l l_1 l_2}B_{l l_a l_b}\begin{pmatrix}l & l_1 & l_2 \\ m & m_1 & m_2\end{pmatrix}\begin{pmatrix}l & l_a & l_b \\ m & m_a & m_b\end{pmatrix}. \tag{C4}$$

Using Eq. (3), (C2), and (C3), we may rewrite Eq. (C4) in a form amenable to rapid computation using filtered

real-space maps and FFTs:

$$\hat{\mathcal{B}}_1 \;=\; \sigma_0^{-2} \sum_{lm} \left\{ \frac{\mathcal{V}_{lm}\mathcal{V}_{lm}^* + 4U_{lm}U_{lm}^* + 2\left(U_{lm}\mathcal{V}_{lm}^* + U_{lm}^*\mathcal{V}_{lm}\right)}{2C_l} \right\}, \tag{C5}$$

$$\mathcal{V}_{lm} \;=\; \int dr\, r^2 \alpha_l(r) B_{lm}^{(2)}(r), \quad B_{lm}^{(2)}(r) = \int d\hat{n}\, Y_{lm}^*(\hat{n}) B^2(r,\hat{n}), \tag{C6}$$

$$U_{lm} \;=\; \int dr\, r^2 \beta_l(r) \int d\hat{n}\, Y_{lm}^*(\hat{n}) A(r,\hat{n}) B(r,\hat{n}). \tag{C7}$$

Written this way, the normalization of the estimator may be computed from a map, and may be efficiently evaluated using FFTs, although the computation of $\mathcal{V}_{lm}$ via numerical integration does introduce a new bottleneck. The operation count for these procedures is $\sim l_{\max}^{5/2} \ln{(l_{\max})} N_{\mathrm{int}}$ where N_{int} is the number of points sampled along the line of sight for the radial integral. The $l_{\max}^{5/2} \ln{(l_{\max})}$ scaling follows from the computational expense of a harmonic transform (in which an FFT is used for the azimuthal Fourier transform piece). In contrast, the direct evaluation of Eq. (44) would require evaluating and summing $\sim l_{\max}^{10}$ terms. The computational savings is then a factor of $l_{\max}^{7.5}/(\ln{(l_{\max})} N_{\mathrm{int}})$, a huge savings for $l_{\max} \sim 10^3$, as is the case for the WMAP and Planck missions. In Ref. [26], it is found that to obtain convergence in the bispectrum estimator itself, $N_{\mathrm{int}} \sim 10^3$ is more than adequate. We thus estimate that use of the real-space RNE will to a computational savings factor of 10^{18} over the harmonic-space estimator.

[1] E. Komatsu et al. (WMAP Collaboration), Astrophys. J. Suppl. Ser. **192**, 18 (2011), arXiv:1001.4538.
[2] D. Salopek and J. Bond, Phys. Rev. D **42**, 3936 (1990).
[3] T. Pyne and S. M. Carroll, Phys. Rev. D **53**, 2920 (1996), arXiv:astro-ph/9510041.
[4] S. Mollerach and S. Matarrese, Phys. Rev. D **56**, 4494 (1997), arXiv:astro-ph/9702234.
[5] S. Mollerach et al., Astrophys.J. **453**, 1 (1995), arXiv:astro-ph/9503115.
[6] D. Munshi, T. Souradeep, and A. A. Starobinsky, Astrophys. J. **454**, 552 (1995), arXiv:astro-ph/9501100.
[7] S. Mollerach, Phys.Lett. **B242**, 158 (1990).
[8] D. H. Lyth, C. Ungarelli, and D. Wands, Phys. Rev. D **67**, 023503 (2003), arXiv:astro-ph/0208055.
[9] D. H. Lyth and D. Wands, Phys.Lett. **B524**, 5 (2002), arXiv:hep-ph/0110002.
[10] C. Gordon and A. Lewis, Phys. Rev. D **67**, 123513 (2003), arXiv:astro-ph/0212248.
[11] C. Cheung, A. L. Fitzpatrick, J. Kaplan, L. Senatore, and P. Creminelli, JHEP **3**, 014 (2008), 0709.0293.
[12] M. Alishahiha, E. Silverstein, and D. Tong, Phys. Rev. D **70**, 123505 (2004), arXiv:hep-th/0404084.
[13] N. Arkani-Hamed et al., JCAP. **0404**, 001 (2004), arXiv:hep-th/0312100.
[14] G. Dvali, A. Gruzinov, and M. Zaldarriaga, Phys. Rev. D **69**, 023505 (2004), arXiv:astro-ph/0303591.
[15] J. Martin, A. Riazuelo, and M. Sakellariadou, Phys. Rev. D **61**, 083518 (2000), arXiv:astro-ph/9904167.
[16] A. Gangui et al., Astrophys. J. **430**, 447 (1994), arXiv:astro-ph/9312033.
[17] A. Gangui, Phys. Rev. D **50**, 3684 (1994), arXiv:astro-ph/9406014.
[18] E. Komatsu and D. N. Spergel, Phys. Rev. D **63**, 063002 (2001), arXiv:astro-ph/0005036.
[19] L. Verde, L.-M. Wang, A. Heavens, and M. Kamionkowski, Mon. Not. R. Astron. Soc. **313**, L141 (2000), astro-ph/9906301.
[20] P. Creminelli and M. Zaldarriaga, JCAP. **0410**, 006 (2004), arXiv:astro-ph/0407059.
[21] K. M. Smith, M. LoVerde, and M. Zaldarriaga, Phys. Rev. Lett. **107**, 191301 (2011), arXiv:1108.1805.
[22] T. L. Smith and M. Kamionkowski (2012), 1203.6654.
[23] D. Babich and M. Zaldarriaga, Phys. Rev. D **D70**, 083005 (2004), arXiv:astro-ph/0408455.
[24] D. Babich, Phys. Rev. D **72**, 043003 (2005), arXiv:astro-ph/0503375.
[25] The Planck Collaboration (2006), arXiv:astro-ph/0604069.
[26] E. Komatsu, D. N. Spergel, and B. D. Wandelt, Astrophys. J. **634**, 14 (2005), arXiv:astro-ph/0305189.
[27] K. M. Smith and M. Zaldarriaga, Mon. Not. R. Astron. Soc. **417**, 2 (2011), astro-ph/0612571.
[28] D. Munshi and A. Heavens, Mon. Not. R. Astron. Soc. **401**, 2406 (2010), arXiv:0904.4478.
[29] K. M. Smith, L. Senatore, and M. Zaldarriaga, JCAP. **0909**, 006 (2009), arXiv:0901.2572.
[30] A. P. S. Yadav and B. D. Wandelt, Phys. Rev. Lett. **100**, 181301 (2008), arXiv:0712.1148.
[31] P. Creminelli, L. Senatore, and M. Zaldarriaga, JCAP. **0703**, 019 (2007), arXiv:astro-ph/0606001.
[32] M. Liguori et al., Phys. Rev. D **76**, 105016 (2007), arXiv:0708.3786.
[33] T. L. Smith, M. Kamionkowski, and B. D. Wandelt, Phys. Rev. D **84**, 063013 (2011), arXiv:1104.0930.
[34] F. Elsner and B. D. Wandelt, Astrophys. J. **724**, 1262 (2010), arXiv:1010.1254.
[35] M. J. White and W. Hu, Astron.Astrophys. **321**, 8 (1997), astro-ph/9609105.
[36] M. J. Rees and D. W. Sciama, Nature (London) **217**, 511 (1968).
[37] R. G. Crittenden and N. Turok, Phys. Rev. Lett. **76**, 575 (1996), arXiv:astro-ph/9510072.
[38] S. P. Boughn and R. Crittenden, New Astron. Rev. **49**, 75 (2005), arXiv:astro-ph/0404470.

[39] T. Giannantonio et al., Phys. Rev. D **74**, 063520 (2006), arXiv:astro-ph/0607572.
[40] S. Ho et al., Phys. Rev. D **78**, 043519 (2008), arXiv:0801.0642.
[41] T. Giannantonio et al., Phys. Rev. D **77**, 123520 (2008), arXiv:0801.4380.
[42] C. L. Francis and J. A. Peacock, Mon. Not. R. Astron. Soc. **406**, 2 (2010), arXiv:0909.2494.
[43] A. P. S. Yadav et al., Astrophys. J. **678**, 578 (2008), arXiv:0711.4933.
[44] U. Seljak and M. Zaldarriaga, Astrophys. J. **469**, 437 (1996), arXiv:astro-ph/9603033.
[45] M. Liguori, S. Matarrese, and L. Moscardini, Astrophys.J. **597**, 57 (2003), astro-ph/0306248.
[46] D. N. Spergel and D. M. Goldberg, Phys.Rev. **D59**, 103001 (1999), astro-ph/9811252.
[47] M. Liguori et al., Phys. Rev. D **76**, 105016 (2007), 0708.3786.
[48] J. M. Mead, A. Lewis, and L. King, Phys. Rev. D **83**, 023507 (2011), arXiv:1009.1549.
[49] E. C. Fieller, Biometrika **24**, 428 (1932).
[50] A. Lewis, A. Challinor, and A. Lasenby, Astrophys. J. pp. 473+ (2000), astro-ph/9911177, URL `http://dx.doi.org/10.1086/309179`.
[51] W. T. Hu, Ph.D. thesis, University of California, Berkeley (1995).
[52] W. Hu, Phys. Rev. D **62**, 043007 (2000), arXiv:astro-ph/0001303.
[53] T. Okamoto and W. Hu, Phys. Rev. D **67**, 083002 (2003), arXiv:astro-ph/0301031.
[54] S. Ho, C. Hirata, N. Padmanabhan, U. Seljak, and N. Bahcall, Phys.Rev. **D78**, 043519 (2008), 0801.0642.
[55] S. Joudaki and M. Kaplinghat (2011), arXiv:1106.0299.
[56] B. M. Schaefer, M. Douspis, and N. Aghanim, Mon. Not. R. Astron. Soc. **397**, 925 (2009), arXiv:0903.4288.
[57] N. Dalal et al., Phys. Rev. D **77**, 123514 (2008), arXiv:0710.4560.
[58] M. Kamionkowski, T. L. Smith, and A. Heavens, Phys. Rev. D **83**, 023007 (2011), arXiv:1010.0251.